CHARACERIZING DRIVER GENE HETEROGENEITY AND CLONAL ORIGIN OF INTRADUCTAL PAPILLARY MUCINOUS NEOPLASMS OF THE PANCREAS

by

Catherine G. Fischer

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

June 2019

ABSTRACT

Intraductal papillary mucinous neoplasms (IPMNs) are precancerous lesions that can progress to invasive pancreatic cancer and a key system in which to study early pancreatic tumorigenesis. We used a combination of multi-region and single-cell targeted next generation sequencing to assess the diversity of somatic driver gene mutations in IPMNs. The resulting data, combined with evolutionary modeling, whole exome sequencing, and *in situ* mutation detection, show that the earliest stages of pancreatic tumorigenesis are characterized by independent clones with distinct early driver gene mutations, thus revealing the polyclonal origin of precancerous pancreatic neoplasms. In addition, multiple mutations in later-occurring driver genes were also common and were frequently localized to unique tumor clones, raising the possibility of convergent evolution of these genetic events in pancreatic tumorigenesis. Collectively, our data demonstrate substantial genetic heterogeneity within IPMN, predominately in IPMNs with low-grade dysplasia. These data also challenge traditional monoclonal tumor origin models and transform our understanding of the evolutionary history of pancreatic neoplasia. Understanding the mechanism underlying polyclonal precancers may reveal new strategies to identify patients at increased risk of developing invasive pancreatic cancer.

Thesis Advisor: Dr. Laura Wood

Thesis Readers: Dr. Laura Wood, Dr. James R. Eshleman

ACKNOWLEDGEMENTS

This work would not have been completed without the invaluable support and contributions from numerous people. First, my advisor Laura Wood provided support in countless ways, and words cannot express how grateful I am for your mentorship. Your brilliance, work ethic, and kindness is inspiring, and I feel incredibly lucky to have been your first graduate student, to see the lab (and your family!) grow. I would also like to thank my thesis committee members, James R. Eshleman, Rachel Karchin, and Victor Velculescu, for the thought-provoking discussions and helpful guidance throughout my graduate studies. I am exceedingly thankful for the scientific and emotional support from the other graduate students in the Wood Lab – Alicia Braxton, Peter Chianchiano, Yea Ji Jeong, Bernat Navarro-Serer, and Ale Trujillo are all incredible scientists and friends, and I was very fortunate to have such an encouraging and fun lab family. In addition, I must thank previous lab members Waki Hosoda, Yuko Kuboki, and Gemma Lionheart for providing scientific expertise, and remarkable friendship. The rest of the Wood Lab, Roberts Lab, GI Pathology Department, and Pathobiology Program also contributed in important ways to this work and to my education. Finally, I would like to thank my parents, Victor and Lois Guerra, my husband, Joe Fischer, and furbabies, Wnt and Illie. Without your unconditional love and unwavering support (even in the form of cuddles), none of my achievements would have been possible. Thank you everyone!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

INTRODUCTION

INTRODUCTION

Pancreatic cancer is expected to be the second leading cause of cancer-related deaths in the U.S. by 2030, with a current 5-year survival rate of only 8%[1,2]. This dismal prognosis is largely due to a lack of early clinical symptoms, with many patients already presenting with advanced disease at initial diagnosis. Therefore, early detection approaches will be critical to improve outcomes in this disease. A great opportunity for early detection is the treatment of premalignant pancreatic lesions, including pancreatic intraepithelial neoplasia (PanIN) and intraductal papillary mucinous neoplasm (IPMN). However, prevention of pancreatic cancer must be balanced with potential overtreatment of low-risk lesions. Recent advances in sequencing technologies have deepened our understanding of the genetic changes that characterize these lesions, which offer new opportunities for screening and early detection. While the majority of PDACs likely arise through PanINs, IPMNs present unique clinical challenges as they are common in the population and are often incidentally identified on routine abdominal imaging[3]. Moreover, their size and resultant early clinical intervention provide critical human tissue samples, which can be used to interrogate early pancreatic neoplasia.

Intraductal Papillary Mucinous Neoplasm (IPMN)

IPMN is a common precursor to pancreatic cancer and the most common cyst of the pancreas. Originally thought to be uncommon, improvements and expanded use of imaging modalities have revealed that nearly 14% of the U.S. adult population harbors a pancreatic cyst[3,4]. IPMNs are grossly visible (>1cm in diameter), mucin-producing neoplastic cysts that arise within the main pancreatic duct or branch ducts[5]. IPMNs were formerly grouped by a three-tier classification system based on dysplasia: low,

intermediate, or high-grade. This classification was revised to a two-tier system where the former intermediate-grade category is now considered low-grade[6]. IPMNs are also categorized by histologic subtype based on the direction of differentiation of the lining epithelium: gastric, intestinal, pancreatobiliary, and oncocytic. Recently, oncocytic-type IPMNs, often referred to as intraductal oncocytic papillary neoplasms (IOPNs), were shown to be unique neoplasms from IPMNs and genetically distinct from the other histologic subtypes[7,8]. Gastric, intestinal and pancreatobiliary-type IPMNs can progress to conventional ductal/tubular carcinomas, while intestinal-type IPMNs can also give rise to colloid carcinomas characterized by extensive stromal mucin accumulation[9,10]. The risk of malignancy associated with an IPMN varies depending on numerous factors (i.e. size, location, grade of dysplasia), with 30-50% of surgically resected IPMNs harboring invasive carcinoma[11–13]. Molecular studies of IPMN have demonstrated that the progression from low-grade IPMN to high-grade IPMN is associated with an accumulation of genetic changes that eventually give rise invasive carcinoma[14]. Therefore, an understanding of the molecular drivers that characterize IPMNs is critical for developing effective early detection strategies for pancreatic cancer.

Molecular Features of IPMN

Numerous studies have identified genetic alterations that play a key role in IPMN tumorigenesis. The most common alterations in IPMN are somatic mutations in the oncogenes *KRAS* and *GNAS*. Mutations in *KRAS* occur in 50-80% of all IPMNs and are thought to be an early event of IPMN development[15–18]. The identification of *GNAS* mutations in IPMN was pioneered by Wu *et al*, which defined a new pathway for pancreatic neoplasia[19]. Since then, *GNAS* alterations have been found in 40-70% of

all IPMNs and remarkably, are not typically found in other pancreatic precursors or in invasive PDAC not associated with an IPMN[18–21]. In association with IPMN, *GNAS* mutations are found in 23-37% of invasive carcinomas[18,21,22]. *GNAS* encodes the alpha-subunit of a stimulatory guanine nucleotide-binding protein, which activates the cyclic-AMP cascade, leading to cell growth and proliferation[23]. *GNAS* mutations are most prevalent in intestinal-type IPMNs, found in 70-100% of these neoplasms[18,19,24,25]. Overall, more than 90% of all IPMNs harbor a *KRAS* and/or *GNAS* mutation, making them important drivers of IPMN development. Moreover, their prevalence in low-grade IPMNs suggests that alterations in these oncogenes may be initiating events in IPMN formation.

Another commonly mutated gene in IPMN is *RNF43*, found in 10-75% of IPMNs[15,22,24]. Alterations in *RNF43* are typically inactivating (i.e. nonsense, frameshift) and accompanied by loss of heterozygosity, implicating *RNF43* as a tumor suppressor gene in IPMN tumorigenesis. The RNF43 protein is a transmembrane E3 ubiquitin ligase that serves as a negative regulator of Wnt signaling, thereby inhibiting cell proliferation[26]. Whole-exome and targeted sequencing studies have identified mutations in *RNF43* in 6-11% of all invasive PDACs, not just those associated with an IPMN[20,27,28].

Inactivation of *CDKN2A*/p16 has been shown to play a role in IPMN progression, particularly during late-stage development. Many studies performing IHC have found p16 loss in IPMNs with high-grade dysplasia (50-100%), and less frequently in low-grade IPMNs (10-51%)[29,30]. However, several next-generation sequencing (NGS) studies have described *CDKN2A* mutations at a much lower prevalence: 0-2% of low-

grade IPMNs and 0-15% of high-grade IPMNs[15,17,24]. This discrepancy may be explained by alternative mechanisms of *CDKN2A* gene silencing. For example, epigenetic silencing by promoter hypermethlyation has been described in 21% of high-grade IPMNs[31]. Additionally, allelic loss of chromosome 9p was found in 18-62% of IPMNs[32,33]. Altogether, these studies indicate that loss of *CDKN2A/*p16 is mediated by the same three mechanisms described in PanIN and is a later event in IPMN tumorigenesis.

Mutations in *TP53* are extremely rare in low-grade IPMNs but appear much more frequently in high-grade IPMNs. Prior to the widespread use of NGS technologies, the literature described variable p53 expression in IPMNs[34]. Studies have found diffuse nuclear p53 labeling in invasive carcinomas but were unable to detect p53 expression in IPMN[35]. Others showed variable p53 staining in 40-50% of high-grade IPMNs[36]. Targeted, massively-parallel sequencing studies have identified *TP53* mutations in 15-20% of high-grade IPMNs and 0-5% of low-grade IPMNs[17,24]. These data suggest that *TP53* mutations are late-occurring alterations and may play a role in the malignant progression of IPMN.

Unlike the previously mentioned tumor suppressor genes, loss of *SMAD4* is mainly confined to invasive carcinomas. Several studies have analyzed the immunohistochemical expression of SMAD4 in IPMNs and invasive cancers[29,37,38]. They all found retained expression of SMAD4 in the vast majority of IPMNs, while typically half of invasive carcinomas show loss of SMAD4. In concordance with the IHC findings, targeted and whole exome sequencing studies also found *SMAD4* mutations to be very rare in IPMN[15,17,24]. However, loss of heterozygosity (LOH) studies using

polymerase-chain reaction (PCR)-based microsatellite analysis found allelic loss of 18q in 22-38% of IPMNs[32,33]. Overall, it seems *SMAD4* inactivation is not involved in early-stages of IPMN development, but important for its transition to invasive carcinoma.

Mutations in several other cancer-related genes have been reported in IPMNs at low prevalence, such as *PIK3CA*, *BRAF*, *PTEN*, and *STK11*[39–43]. Mutations in the oncogenes *PIK3CA* and *BRAF* have also been reported to be drivers in invasive PDAC and many other cancer types[20]. Early studies have reported loss-of-function mutations in *PTEN* at relatively low frequencies in invasive PDAC; however, more recent studies have found loss of at least one copy of the *PTEN* gene can help drive malignant progression of both human and mouse PDACs[44,45]. Peutz-Jeghers syndrome patients have an elevated risk of pancreatic malignancy, and commonly harbor germline mutations in the tumor suppressor gene *STK11*[46,47]. Several studies have demonstrated LOH at the *STK11* locus in sporadic PDACs and other cancers of the gastrointestinal tract, breast, and ovaries[48–51].

Early detection of IPMN

As mentioned previously, IPMN is a common precursor of PDAC and therefore represents a key target for early detection approaches. Imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), and endoscopic ultrasound (EUS) are commonly used to detect lesions in the pancreas. In a multi-institution study conducted by Canto *et al*, these imaging methods were used to screen 225 asymptomatic, high-risk individuals (HRI)[52]. They found that 42% of HRIs had at least one pancreatic mass or a dilated pancreatic duct. Among these, proven or suspected lesions were later identified in 92% of patients. These imaging modalities can

also be augmented to target specific structures and molecules. For example, Neesse *et al* designed a fluorochrome that specifically targets claudin-4[53], a protein known to be upregulated in pancreatic neoplasia[54,55].

While these imaging-based approaches are useful to detect pancreatic cysts, they may not reliably differentiate cyst type or important histological features (i.e. grade of dysplasia), which can better predict likelihood of progression. This distinction is clinically important because pancreatic cysts represent a diverse group of lesions, some of which are low-risk while others progress to invasive carcinoma. As a result, a more reliable determination of precursors with a higher malignant potential will be critical. Many studies have demonstrated the value of collecting cyst fluid by EUS-fine needle aspiration from patients diagnosed with pancreatic cysts. Several reports highlight the importance of cytological evaluation of cyst fluid for atypical epithelial cells, which can serve as a predictor of malignancy[56–58]. Others have reported on the diagnostic value of biochemical markers in cyst-fluid for differentiating likely benign, serous cysts from mucinous cysts which have greater risk of malignancy[59–61]. Furthermore, two independent studies used targeted NGS to analyze cyst-fluid[62,63]. The investigators used a combination of molecular markers to categorize a cyst as IPMN with 76-100% sensitivity and 84-96% specificity. This approach identified IPMNs with high-grade dysplasia or invasive carcinoma with 88% sensitivity and 69-97% specificity. These results highlight the ability of cyst fluid sequencing to preoperatively determine cyst type as well as predict grade of dysplasia in premalignant cysts. Another marker that could be used to differentiate grade of dysplasia in IPMN is telomere fusion, which frequently occurs in critically short telomeres. Hata *et al* developed a real-time quantitative PCR

method to detect telomere fusion in cyst-fluid[64]. They detected telomere fusions in 27% of high-grade IPMNs, but not in low-grade IPMNs. Additionally, several studies have used cyst fluid to identify other molecular changes such as telomerase activity and microRNA levels[65,66]. As a result from these findings and others, several institutions are implementing NGS-based molecular tests using cyst fluid to aid in the clinical evaluation and diagnosis of pancreatic cysts.

While these studies provide important data for the early detection of IPMN, they are limited by sampling only cyst-fluid. Felsenstein *et al* found that a substantial portion of PDACs with co-occurring IPMNs are unrelated[22]; therefore, analysis of cyst fluid may not detect the true precursor of the cancer. A promising approach is genetic analysis of secretin-stimulated pancreatic juice collected from the duodenum. Yu *et al* developed a digital NGS approach to detect low-abundance mutations in pancreatic juice samples[67]. In two cases of high-risk individuals, digital NGS was able to detect *SMAD4* or *TP53* mutations more than one year before their pancreatic cancer diagnosis. Finally, Suenaga *et al* performed digital NGS using a targeted 12-gene panel to evaluate mutation concentrations in pancreatic juice samples[68]. Mutation concentrations in genes other than *KRAS*/*GNAS* were higher in patients with PDAC or high-grade precursors relative to all other subjects. Consistent with previous studies and the molecular progression of IPMN, this pancreatic juice analysis found several predictors of pancreatic cancer or a high-grade precursor: presence of *SMAD4* mutations, high *SMAD4/TP53* mutation concentrations, and high overall mutation concentrations. While the lack of these mutations does not guarantee absence of disease, these results highlight the potential value of a pancreatic juice NGS screening

test for patients undergoing pancreatic evaluation. Finally, analysis of molecules and cells in the blood also does not require sampling of a specific lesion in the pancreas and thus is an alternative approach to early detection. For example, some groups have reported analysis of circulating epithelial cells (CECs) in the bloodstream in patients with pancreatic cancer as well as IPMN[69,70]. This method has been used to detect CECs in patients with cystic lesions prior to the clinical diagnosis of invasive PDAC. Numerous serum markers have also been suggested as potential biomarkers for early detection of pancreatic neoplasia, but these remain to be systematically evaluated[71].

Genetic Heterogeneity in IPMN

Genetic heterogeneity of driver gene alterations in IPMN have recently been described due to increased sensitivity of NGS approaches. Wu *et al* used targeted sequencing to analyze cyst fluid from 19 patients[19]. They found 11% of IPMNs contained two different *KRAS* mutations, 2% contained three different *KRAS* mutations, and 4% contained two different *GNAS* mutations. Additionally, Felsenstein *et al* microdissected epithelium from two distinct regions of IPMN, and subsequently performed deep targeted sequencing of pancreatic driver genes[22]. These studies reported that 23% of IPMNs had multiple *KRAS* and/or *GNAS* mutations, and remarkably one of these IPMNs contained four unique *KRAS* mutations.

In the future, more comprehensive analyses of the genetic heterogeneity in pancreatic cancer precursor lesions can further elucidate their clonal evolution and neoplastic progression. Additionally, the identification and validation of molecular markers that can reliably distinguish low-risk lesions from lesions with a high risk of malignant progression will be fundamental for effective early detection approaches.

These promising approaches require a deep understanding of the molecular alterations that occur during pancreatic tumorigenesis.

My thesis work focused on characterizing genetic heterogeneity within IPMNs and exploring the evolutionary history of early pancreatic tumorigenesis. This work can be divided into two sections. Chapter 2 focuses on a comprehensive multi-region targeted sequencing study, which analyzed 227 neoplastic samples from 20 cases of surgically resected IPMN. This targeted sequencing data was further supported by evolutionary modeling, whole exome sequencing, and *in-situ* mutation detection. Chapter 3 supplements the previously mentioned work by using the novel approach of single-cell sequencing of IPMNs. Altogether, this thesis work utilized multiple sequencing-based approaches to reveal extensive driver gene heterogeneity and polyclonal origin of IPMN. These findings provide novel insights into the initiation and progression of pancreatic precancers, and advance our understanding of the origins of invasive pancreatic cancer.

The following represent unanswered questions that require further investigation:

- Why do these early pancreatic precancers contain multiple, independent clones?
- What mechanisms support the growth of polyclonal pancreatic precancers?
- What underlying features/phenomena lead to polyclonal origin in early pancreatic precancers?

CHAPTER 2:

Polyclonal origin defines early pancreatic tumorigenesis

Rationale

While the majority of PDACs likely arise through PanINs, IPMNs present unique clinical challenges as they are common in the population and are often incidentally identified on routine abdominal imaging[3]. Moreover, their size and resultant early clinical intervention provide critical human tissue samples, which can be used to interrogate early pancreatic neoplasia. Molecular studies of IPMN suggest that the progression from low-grade to high-grade dysplasia is associated with an accumulation of genetic alterations in oncogenes and tumor suppressor genes, eventually leading to the development of invasive PDAC. Previous studies have characterized several key driver gene alterations in IPMNs: mutations in the oncogenes *KRAS* and *GNAS* are thought to be the earliest driver gene alterations, while mutations in tumor suppressor genes such as *RNF43*, *CDKN2A*, and *TP53* occur later[72]. With the exception of *GNAS*, these driver genes are also commonly found in invasive PDACs occurring in the absence of IPMN[73]. Most studies of malignant progression in IPMNs analyzed a single tissue section from each lesion across many patients, allowing comparison of the genetic alterations among different IPMNs based on grade of dysplasia or other clinical and pathological features[15,17,21,24,74]. Although such studies allow estimation of mutation prevalence in specific IPMN subgroups, they do not provide comprehensive data about the diversity of genetic alterations within a single lesion. Recent studies that analyzed more than one section from each IPMN suggest the potential for substantial genetic heterogeneity within these neoplasms; however, analyses systematically interrogating this genetic heterogeneity are currently lacking. Such studies can provide important insights into the evolutionary history of early pancreatic tumorigenesis.

Importantly, recent data have shown a lack of genetic heterogeneity with respect to driver gene alterations in invasive PDAC, raising the possibility that unique evolutionary processes govern the earlier stages of tumorigenesis in the pancreas[75,76]. As our current understanding of pancreatic tumor evolution has mostly been inferred from sequencing advanced cancers, there is a critical need for analyses of *bona fide* precancers to directly describe initiation and progression in pancreatic neoplasia. The importance of studying precancers has been exemplified in other tumor types, such as esophageal carcinogenesis[77,78]. These studies reshaped our understanding of the origin and dynamics of tumorigenesis in these organs and have significant biologic and therapeutic implications.

To explore the evolutionary history of early pancreatic tumorigenesis, we performed multi-region targeted next generation sequencing on surgically resected IPMNs. These data provide novel insights into IPMN progression, which will advance our understanding of the origins of invasive pancreatic cancer.

Materials and Methods

***Specimen acquisition*.** This study was approved by the Institutional Review Board of the Johns Hopkins Hospital. We retrospectively collected multi-region samples from 20 patients diagnosed with IPMN who had undergone surgical resection at the Johns Hopkins Hospital between 2008 and 2015. Slides from every available formalin-fixed paraffin embedded (FFPE) block from each IPMN were stained with hematoxylin-and-eosin (H&E). An H&E slide from each block was reviewed by an expert pancreatic pathologist to confirm diagnosis and determine regions for microdissection. One H&E

slide from each matched normal (duodenum or spleen) was also reviewed by an expert pathologist prior to DNA extraction.

***Laser capture microdissection*.** Five 10µm serial tissue sections from FFPE blocks were cut onto membrane slides (Zeiss MembranSlide 1.0 PEN). Deparaffinization was performed in fresh xylenes for 2 min, followed by 100% ethanol for 2 min, 95% ethanol for 2 min, and 70% ethanol for 2 min. Subsequently, the slides were stained by crystal-violet (Sigma Aldrich; diluted 1:4 in 70% ethanol) for 30 sec and dehydrated by ascending ethanol solutions. The stained slides were microdissected within 1 h. Regions of IPMN epithelium were identified under the microscope and microdissected for enrichment of neoplastic cellularity on a Leica LMD7000 instrument. Microdissected tissues were collected into 0.5ml Lo-bind tubes (Eppendorf) and immediately processed for subsequent DNA extraction. For matched-normal samples, five 5µm sections were cut onto regular slides and tissue was scraped off using a sterile razor blade (Personna).

***DNA extraction and quantification*.** DNA was extracted from each sample using the QIAamp DNA FFPE Tissue kit (Qiagen) and the MagMAX FFPE Isolation kit (Applied Biosystems). Briefly, 180µl ATL buffer and 20µl proteinase K was added to the 0.5ml Lo-bind tube containing tissue and subsequently incubated on an agitating thermomixer for 24h at 56C. After this period, 2µl of MagMAX protease was added and incubated on a stationary heat block for 1h at 60C, followed by incubation for 30min at 80C. Next, 200µl of AL buffer and 1µg of carrier RNA were added to the sample and incubated for 20min at 70C. Finally, ethanol precipitation and column elution were conducted using the QIAamp DNA FFPE Tissue kit according to the manufacturer's instructions. DNA

concentration was measured using the Qubit dsDNA High Sensitivity Assay kit (Invitrogen) according to the manufacturer's protocol on a Qubit 2.0 fluorometer (Invitrogen). DNA was stored at -20C until library preparation.

***Targeted sequencing and analysis*.** A targeted sequencing approach analyzed the entire coding regions of 15 known driver genes in IPMN tumorigenesis (*KRAS*, *GNAS*, *BRAF*, *RNF43*, *CDKN2A*, *PIK3CA*, *PTEN*, *APC*, *CTNNB1*, *MAP2K4*, *STK11*, *ATM*, *TP53*, *TGFBR2*, *SMAD4*). Library preparation was performed using the Agilent SureSelect^{XT} Target Enrichment System (Agilent) following the manufacturer's instructions with the following modifications: genomic DNA was sheared to 200-250bp fragments using a Covaris M220 Focused-ultrasonicator (Covaris). Additionally, both pre- and post-capture PCR cycle number was increased by 0, 1, or 2 for DNA inputs of 200ng, 50-100ng, or 25-50ng, respectively. The barcoded libraries were sequenced using an Illumina MiSeq system (Illumina), generating 300 base pairs (2x150bp reads) per fragment. The average distinct sequencing depth was 570x per sample. Sequence alignment files were processed according to Genome Analysis Toolkit (GATK) best practices[79,80]. Sequences were aligned using Burrow-Wheeler Aligner (BWA)[81] to the human reference genome (UCSC hg19). Duplicate reads were marked with Picard Tools[82]. Somatic mutations were identified by processing matched normal and tumor reads with MuTect (single nucleotide variants) and MuTect 2 (indels)[83]. We identified candidate mutations that were altered in ≥5% of distinct reads with coverage >100X. All candidate mutations were confirmed or rejected following visual inspection in Integrated Genome Viewer (IGV)[84]. To not over-estimate genetic heterogeneity, every mutation called in at least one sample within a given IPMN was inspected at that locus in the

other samples to ensure sufficient coverage of >100x - thus, minimizing false negative calls. Additionally, previous modeling on the probability of detecting mutations at a given depth of coverage ensures an extremely low likelihood of representing false negatives given DNA input and coverage[85].

***Somatic mutation evolutionary reconstruction*.** We reconstructed somatic mutation evolutionary trees based on the neoplastic cell fractions (NCFs) of the mutations we identified, using SCHISM software[86]. Mutation NCFs were calculated using VAF, estimated tumor purity (p), tumor copy number (CN_T) at the mutation site, germline copy number (CN_N) at mutation site, and multiplicity (m), i.e. estimated number of mutant alleles.

$$\boldsymbol{NCF = \frac{1}{m}\,\frac{p\,CN_T + (1-p)CN_N}{pVAF}}$$

Tumor purity and copy number are difficult to estimate from a small targeted panel. Rather than assuming a single value for purity, on which both CN_T and NCF depend, we tried purity values in the range 0.5 to 0.9, in increments of 0.1. CN_T and *m* were estimated at each purity level. Tumor integer copy number was estimated by scaling the tumor-to-normal read ratios to consider purity. In the absence of allele-specific copy number information, *m* was set to 1 unless VAF>0.8. For mutations with VAF>0.8, *m* was set to the CN_T if $CN_T \leq 2$, and CN_T-1 if $CN_T > 2$.

For each IPMN, SCHISM was used to estimate an optimal evolutionary tree at each purity level. We chose the optimal tree based on the highest fitness value calculated by SCHISM. When SCHISM identified multiple trees with the same highest

fitness value, we present all trees in Supplementary Figure 1. If SCHISM could not infer a tree with a fitness >0.1, we excluded the case from the evolutionary analysis. SCHISM was originally designed to model only tumors of monoclonal origin, and we extended it to handle topologies with multiple originating clones.

***Whole-exome sequencing and analysis*.** Human exome capture was performed using Agilent's SureSelect Human All Exon 50Mb Kit 5.0 (Agilent). The captured libraries were sequenced with an Illumina XTEN system, generating 300 base pairs (2x150bp reads) per fragment. The average distinct sequencing depth was 104x per sample. After Illumina sequencing, all produced FASTQ reads were quality-checked and trimmed with FastQC (version 0.11.2) and Trimmomatic (version 0.33)[87]. Sequencing reads were aligned to human genome version hg19 with the BWA MEM software for both tumor and normal samples[81]. PCR duplications were marked with Picard software (version 1.103)[82]. The BAM files were locally realigned, and the base quality scores were recalibrated with GATK (version 3.1)[79]. MuTect (version 1.1.6) was used to call single nucleotide variants with default parameters[83]. Only SNVs that were classified as "KEEP" by MuTect and >10% VAF were used for downstream analysis. Indels were found using Strelka (version 1.0.14) with default parameters[88]. We identified candidate mutations that were altered in ≥10% of distinct reads. All candidate mutations were confirmed or rejected following visual inspection in Integrated Genome Viewer (IGV)[84].

***Single-molecule RNA in situ hybridization*.** The BaseScope assays (Advanced Cell Diagnostics, Inc., Newark, CA) were developed to achieve point mutation-specific detection of the *KRAS* transcripts. The BaseScope assay is based on the RNAscope

technology[89] in which paired double-Z oligonucleotide probes were designed against target RNA using custom software, but BaseScope uses an additional signal amplification step and requires only one “double Z” (1 ZZ) probe pair for single-molecule detection. The 1-ZZ probe BA-Hs-KRAS-G12V was designed to target *KRAS* G12V (nt. G35T) mRNA, the 1-ZZ probe BA-Hs-KRAS-G12D was designed to target *KRAS* G12D (nt. G35A) mRNA, and the 1-ZZ probe BA-Hs-KRAS-G12R was designed to target *KRAS* G12R (nt. G34C) mRNA. All the above probes were validated using KRAS control cell lines (G12V: SW620; G12D: SNU-C2B; G12R: PSN1) for probe specificity. The BaseScope™ Reagent Kit – RED (Advanced Cell Diagnostics) was used for sample pretreatment, hybridization and signal development according to the manufacturer’s instructions. FFPE tissue section samples were prepared according to manufacturer’s recommendations. Each FFPE sample was quality controlled for RNA integrity with a BaseScope 1-ZZ probe specific to the housekeeping gene Hs-PPIB. Negative control background staining was evaluated using a 1-ZZ probe specific to the bacterial *dapB* gene.

***Estimating genetic heterogeneity*.** Genetic heterogeneity was quantitatively measured using Jaccard similarity coefficients. The Jaccard similarity coefficient is defined as the ratio of shared variants to all variants (shared + discordant) between two samples. We calculated Jaccard similarity coefficients (based on *KRAS* and *GNAS* mutations only) between all samples for each IPMN. Next, we calculated the average Jaccard similarity coefficient for each IPMN. Additionally, the average coefficient of high-grade IPMNs was recalculated after excluding low-grade samples. The average Jaccard similarity coefficient was plotted for all three groups (low-grade IPMNs, high-grade IPMNs (whole

lesion analysis), and high-grade IPMNs (high-grade only analysis)) using ggplot2 function in R[90] and significant differences between groups were determined by Mann-Whitney *U* test.

***Software*.** OncoPrinter software available through Memorial Sloan Kettering Cancer Center's cBioPortal was used to generate Figure 1. SCHISM software is available for non-profit use at https://karchinlab.org/apps/appSchism.html.

Results

***Clinical and pathological features*.** This study included 227 neoplastic samples from 20 surgically resected IPMNs (Table 2.1). Of the 20 IPMNs analyzed, ten had low-grade dysplasia and ten had high-grade dysplasia. Of the ten high-grade IPMNs, three had co-occurring pancreatic ductal adenocarcinoma (PDAC), one of which occurred in the same blocks as the IPMN. The size of IPMNs in our cohort ranged from 1.6 cm to 6.5 cm, resulting in 6 to 24 FFPE blocks per case. For each case, IPMN epithelium was microdissected from every FFPE block. Importantly, because IPMNs are classified based on the highest grade of dysplasia present in the lesion, high-grade IPMNs will often also contain areas with low-grade dysplasia, even in the same FFPE block. Therefore, areas of low-grade dysplasia were separately microdissected from the areas of high-grade dysplasia. We refer to each FFPE block as an analyzed "region" from which we isolated one or two DNA "samples" based on the grade(s) of dysplasia present within the FFPE block. Of the ten high-grade IPMNs, seven also contained low-grade dysplasia – resulting in more total samples than regions in these cases. We microdissected a total of 117 low-grade samples, 105 high-grade samples, and five invasive PDAC samples. Of the 117 low-grade samples, 42 were from IPMNs that were

overall classified as high-grade. Matched normal DNA for each case from duodenum or spleen was also analyzed using the same sequencing pipeline and used as a germline comparator.

Multi-region sequencing reveals striking driver gene heterogeneity in IPMN. Targeted next generation sequencing (mean distinct coverage >500x) was used to interrogate the entire coding regions of 15 well characterized driver genes in pancreatic tumorigenesis. Based on previous literature that characterized these driver gene mutations, we consider mutations in *KRAS*, *GNAS*, and *BRAF* to be "early drivers," while we consider mutations in the other genes interrogated by our panel to be "later-occurring drivers"[17,72,74]. Mutations (including single nucleotide variants and small insertions/deletions) were identified in *KRAS*, *GNAS*, *BRAF*, *RNF43*, *CDKN2A*, *PIK3CA*, *CTNNB1*, *STK11*, *ATM*, *APC*, and *TP53* (Table 2.1). The total number of somatic driver gene mutations called in each analyzed IPMN ranged from 2 to 9 (Table 2.1). The most commonly mutated genes were *KRAS* and *GNAS* – all IPMNs had at least one *KRAS* mutation, and 17 had at least one *GNAS* mutation (Figures 2.1A and 2.1B). Of the 20 IPMNs, 15 had multiple mutations called in the same driver gene.

This sequencing approach identified mutations that were present in all regions of an IPMN (hereafter referred to as "homogeneous mutations"), as well as mutations that were only present in a subset of the analyzed regions (hereafter referred to as "heterogeneous mutations"). On average, a greater proportion of mutations were heterogeneous in low-grade IPMNs, relative to high-grade (Figure 2.1C). Of the ten low-grade IPMNs, three did not contain any homogeneous mutations, and nine contained at least one heterogeneous mutation (Figure 2.1C). In order to more precisely characterize

the molecular features of IPMN epithelium with different grades of dysplasia, we analyzed IPMNs classified as high-grade in two different ways: (1) we considered both the low-grade and high-grade samples for each IPMN, which we call “whole lesion analysis”, and (2) we considered only the high-grade samples, which we call “high-grade only analysis.” Using whole lesion analysis, 57% of mutations were heterogeneous in high-grade IPMNs. However, this proportion decreased to 36% in the high-grade only analysis, suggesting that high-grade samples were less heterogeneous than low-grade samples within the same IPMN. (Figure 2.1C). For example, IPC15 appeared to contain only heterogeneous mutations via whole lesion analysis; yet, less than half of the mutations were heterogeneous after excluding low-grade samples. These data demonstrate a striking degree of driver gene heterogeneity in IPMNs, especially those classified as low-grade, indicating that the identification of a driver mutation in one region of an IPMN does not necessarily imply its occurrence throughout the entire lesion.

Early driver genetic heterogeneity is pervasive in IPMNs with low-grade dysplasia. All of the IPMNs in our cohort harbored mutations in *GNAS* and/or *KRAS*, and we identified 11 with multiple mutations in these early driver genes**.** These 11 IPMNs all had multiple mutations in *KRAS*, including seven with two different *KRAS* mutations, two with three different *KRAS* mutations, and two with four different *KRAS* mutations. In every sample that contained multiple *KRAS* mutations, each mutation was found on distinct sequencing reads and therefore did not occur on the same allele. Of these 11 IPMNs with multiple *KRAS* mutations, six also had multiple *GNAS* mutations. Strikingly, IPC02 and IPC08 had five and six different total mutations in *KRAS* and

GNAS, respectively (Figures 2.1A and 2.1B). Of the ten low-grade IPMNs, seven harbored multiple mutations in *GNAS* and/or *KRAS*, while this was true for just one high-grade IPMN when performing high-grade only analysis (Fisher's exact test; p=0.02). We quantitatively measured genetic heterogeneity with respect to *KRAS* and *GNAS* by calculating Jaccard similarity coefficients[75,76] (Table 2.2). When using whole lesion analysis, we found that low-grade IPMNs were more genetically heterogeneous than high-grade IPMNs with respect to *KRAS* and *GNAS* (0.63 v. 0.87; Mann-Whitney *U* test, p=0.070). Moreover, when using high-grade only analysis, high-grade IPMNs were significantly less genetically heterogeneous that low-grade IPMNs with respect to these early driver gene alterations (0.63 v. 0.92; Mann-Whitney *U* test, p=0.019) (Figure 2.2A). Altogether, these data reveal genetic heterogeneity with respect to early drivers is significantly more prevalent in low-grade IPMNs.

Evolutionary modeling suggests polyclonal origin of early pancreatic tumorigenesis. Evolutionary trees, built with SCHISM[86] (see Methods), suggested that several IPMNs were polyclonal in origin – having initiated from more than one separate and independent clone (Figure 2.2). For example, we identified three different *KRAS* mutations (*KRAS* p.G12D, p.G12R, p.G12V) in IPC03, each marking a distinct clone (Figure 2.2B). In another example, whole lesion analysis of IPC15 inferred it originated from three clones, each marked by a different early driver (*KRAS* p.G12V, *KRAS* p.G12D, *GNAS* p.R201C) (Figure 2.2C). Interestingly, the clone containing *KRAS* p.G12D and *GNAS* p.R201S was absent in the high-grade samples, demonstrating this clone was limited to the low-grade dysplasia. Strikingly, whole lesion analysis of IPC14 suggested this IPMN initiated from two clones with either a *GNAS* p.R201H or *KRAS*

p.G12R mutation – however, high-grade only analysis revealed that the high-grade neoplasia in IPC14 initiated from a single clone (*GNAS* p.R201H) (Figure 2.2D). Altogether, some high-grade IPMNs appear polyclonal in origin, but IPC15 is the only high-grade IPMN that retains polyclonality after excluding low-grade samples from the analysis.

Whole exome sequencing confirms independent origin of clones with discordant mutations in early driver genes. While our targeted sequencing data coupled with evolutionary modeling suggests polyclonal origin for several IPMNs in our cohort, it is also possible that these multiple *KRAS* mutations occurred as divergent later events in a monoclonal IPMN initiated by an alteration not examined by our targeted panel. In order to more definitively characterize the clonal origin of the analyzed IPMNs, we performed whole exome sequencing (WES) on two regions each from four IPMNs (IPC01, IPC08, IPC12, and IPC14). For each IPMN analyzed by WES, we chose two regions with mutually exclusive early driver gene mutations based on our targeted sequencing. For example, samples T5 HG and T8 LG from IPC14 were chosen for WES because T5 HG had a *GNAS* p.R201H mutation, while T8 LG had a *KRAS* p.G12R and *BRAF* p.E26D (Figure 2.2D). On average, we found ~30 coding somatic mutations per IPMN region and, importantly, we did not identify any shared alterations between the two regions (Tables 2.3-2.6). Additionally, all mutations called using our targeted sequencing approach at >10% VAF were also found via WES at similar frequencies, further validating the accuracy of our multi-region targeted sequencing. Overall, this comprehensive approach has corroborated our evolutionary modeling and confirms these IPMNs originated from more than one distinct clone – providing, for the

first time, strong evidence for polyclonal origin of these neoplasms. If these results are extrapolated to all IPMNs in our cohort harboring multiple early driver gene mutations, this suggests 60% of the analyzed IPMNs are polyclonal in origin.

RNA in-situ hybridization confirms spatial separation of mutant KRAS subclones. The targeted sequencing and evolutionary modeling in the current study suggest that multiple *KRAS* mutations observed within a single IPMN occurred in different clones; however, these data cannot resolve whether the cells containing discordant *KRAS* mutations are intermixed or spatially separated. To determine the spatial location of cells with discordant *KRAS* mutations we employed BaseScope, which uses RNA *in situ* hybridization to specifically detect hotspot alterations in *KRAS* (see Methods). We analyzed one region each from five IPMNs by probing for *KRAS* p.G12V, p.G12D, and p.G12R, which were validated using *KRAS* mutant cell lines. We chose regions to analyze based on several factors, including age of tissue block, amount of IPMN epithelium within block, and identification of *KRAS* mutations targeted by BaseScope probes. Interestingly, in each region that contained multiple *KRAS* mutations, we identified clones that were spatially separated (Figure 2.3, representative case). For example, IPC09 B was a low-grade region with three distinct areas of IPMN epithelium, which were pooled for sequencing (Figure 2.3A). Targeted sequencing indicated the neoplastic cells in this region had *KRAS* p.G12D and p.G12V mutations. The IPMN cells in areas 1 and 2 expressed *KRAS* p.G12D, but not p.G12V (Figures 2.3B-2.3E). Conversely, the IPMN cells in area 3 expressed *KRAS* p.G12V, but not p.G12D (Figure 2.3F and 2.3G). Overall, while areas 1 and 2 were spatially separated, they expressed the same *KRAS* mutation and therefore are likely of the same originating clone. Area 3

was spatially separated from areas 1 and 2 and expressed a different *KRAS* mutation, suggesting this region likely originated from a distinct clone (Figure 2.3H). Importantly, the results of the BaseScope assay for each analyzed IPMN region were concordant with our targeted sequencing with respect to the *KRAS* mutations identified and the relative proportions of cells containing each mutation.

Convergent evolution and heterogeneity of later-occurring driver genes in IPMNs with high-grade dysplasia. In addition to heterogeneity with respect to early driver gene mutations, we also identified genetic heterogeneity with respect to later-occurring driver gene mutations in several IPMNs in our cohort. Multiple *RNF43* mutations were identified in five IPMNs from our cohort. Four IPMNs (IPC07, IPC18, IPC19, and IPC20) had two different *RNF43* mutations, while IPC16 had four different *RNF43* mutations (Figure 2.4A). The four *RNF43* mutations identified in IPC16 were either nonsense or frameshift and interestingly, all four were identified in distinct regions. One *RNF43* mutation (p.R49Sfs*3) was present in both the low-grade and high-grade dysplasia from a single region. A second *RNF43* mutation (p.R132*) was only found in a different region with low-grade dysplasia. The remaining two *RNF43* mutations (p.Q152*and T76Nfs*13) were found in separate regions with only high-grade dysplasia. These data demonstrate multiple, distinct mutations occurring in the *RNF43* gene, suggesting there is a specific selection for these mutations in a subset of IPMNs. We also observed multiple mutations in the *TP53* gene. One high-grade IPMN (IPC20) had four different *TP53* mutations (Figure 2.4B). One *TP53* missense mutation (p.V172D) was present in every region, while the other *TP53* missense mutations (p.P152L, p.D208A, and p.R248Q) were present in only one to three regions each.

While our data suggest low-grade IPMNs are significantly more heterogeneous with respect to early driver mutations, heterogeneity amongst later-occurring driver mutations is prominent in high-grade IPMNs. Of the five IPMNs harboring multiple *RNF43* mutations, four were high-grade IPMNs. Interestingly, while early driver heterogeneity was less prevalent in high-grade IPMNs relative to low-grade, this was not true for later-occurring drivers (Figure 2.4C and 2.4D). Heterogeneity with respect to these later-occurring driver mutations seems to be a feature of high-grade IPMNs.

Mutations in less commonly mutated driver genes also occurred in our cohort. Four IPMNs (IPC06, IPC08, IPC14, and IPC17) had mutations in *BRAF*. IPMNs IPC06 and IPC08 had the same in-frame deletion of *BRAF* (p.N486_P490del), which has previously been reported in IPMNs and numerous other cancers[91,92] and, interestingly, has been shown to confer sensitivity to BRAF inhibitors[93]. In IPC08, we identified two regions which lacked a *KRAS* mutation, yet harbored this *BRAF* deletion (Figure 2.1). Two IPMNs (IPC03 and IPC17) had missense mutations in *CDKN2A*. Two IPMNs (IPC07 and IPC09) had mutations in oncogenic hotspots of *PIK3CA* (p.N1044K and p.H1047R). In IPC09, every region harbored *PIK3CA* p.H1047R and KRAS p.G12V at similar NCF; therefore, we could not determine which mutation occurred first, or if these mutations represent distinct clones. One IPMN (IPC06) had a mutation in an oncogenic hotspot of *CTNNB1*, and one IPMN (IPC19) had a missense *APC* mutation. One IPMN (IPC10) had a missense mutation in *STK11*. Three IPMNs (IPC08, IPC12, and IPC14) had mutations in *ATM*, one of which (IPC12) was an inactivating frameshift insertion. None of the IPMNs in our cohort had mutations in *SMAD4*.

Clinical implications of genetic heterogeneity in IPMN. An independent study analyzed somatic mutations in cyst fluid from patients diagnosed with IPMN prior to surgical resection (*manuscript under review*). Seven cases from this study overlapped with our cohort (IPC03, IPC04, IPC06, IPC08, IPC12, IPC18, and IPC20). We compared driver gene mutations identified in DNA collected from cyst fluid to our multi-region targeted sequencing. For all cases, mutations found in the cyst fluid were also found via our multi-region sequencing approach except for those at very low VAF in the cyst fluid analysis (range 0.00076-0.034 VAF). In all cases, there were mutations identified by our multi-region sequencing approach that were not identified in the cyst fluid. For example, both approaches found *KRAS* p.G12V in IPC03; yet, two additional *KRAS* mutations (p.G12D and p.G12R), and a *GNAS* mutation (p.R201H) were not identified in the cyst fluid analysis (Table 2.7). In general, the mutations not identified in the cyst fluid were present in only a subset of regions within a given IPMN.

Discussion

Because IPMNs are often diagnosed on imaging studies, they can be resected prior to the development of invasive PDAC – as such, IPMNs represent an ideal system in which to perform comprehensive analyses of precancerous pancreatic neoplasia. In this study, we performed multi-region targeted sequencing on 227 neoplastic samples from 20 IPMNs, using meticulous isolation of epithelium via LCM from 6 to 24 regions per IPMN. Thus, our study represents one of the most comprehensive genetic analyses of driver gene mutations in IPMNs to date, and to our knowledge, the only study to interrogate such mutations throughout entire IPMNs. The resulting data strongly

suggest that these precancers are derived from multiple independent clones, transforming our understanding of the origins of pancreatic tumorigenesis.

Previous studies characterizing the genetic alterations that drive the progression of IPMN have typically analyzed a single area from each neoplasm. However, a single area may not harbor the full diversity of driver gene alterations within a single IPMN. More recent studies have suggested the potential for significant genetic heterogeneity in IPMNs – however, these studies did not comprehensively identify the genetic alterations across an entire IPMN[19,22]. Here, we demonstrate that the vast majority of analyzed IPMNs harbor multiple mutations within the same driver gene. In addition, we report that many IPMNs do not have any homogeneous mutations across the entire lesion, demonstrating that single-region sequencing approaches are not adequate to accurately identify the complete repertoire of driver gene mutations within an IPMN. Furthermore, the genes that harbor multiple mutations in IPMNs differ depending on the grade of dysplasia and thus the stage of neoplastic progression. We found IPMNs with low-grade dysplasia have significantly more heterogeneity with respect to the early drivers *KRAS* and *GNAS*, compared to IPMNs with high-grade dysplasia. These data demonstrate that early pancreatic tumorigenesis is characterized by heterogeneity in initiating driver genes, while progression to high-grade dysplasia leads to decreased heterogeneity in these early driver genes. This likely represents selection and expansion of a single clone after acquisition of additional driver gene alterations, thus eliminating prior clonal diversity. This finding is consistent with previous studies which have demonstrated a lack of driver gene heterogeneity in invasive PDACs and their metastases[75,76],

suggesting that this heterogeneity and clonal selection are limited to precancerous pancreatic neoplasms.

Using our multi-region sequencing approach, coupled with evolutionary modeling, we show that a substantial proportion of IPMNs in our study evolved from multiple distinct clones. This polyclonal origin is typically marked by multiple *KRAS* mutations, yet clones can also originate from *GNAS* mutations, albeit less frequently. Intriguingly, one of the analyzed IPMNs harbored two independent clones, one initiated by a mutation in *KRAS* and one by a mutation in *GNAS*. This demonstrates that independent clones can be initiated by mutations in different genes, suggesting that the presence of multiple mutations in a single early driver gene probably underestimates polyclonality. Importantly, WES of IPMN regions with discordant early driver mutations did not identify any shared mutations across the entire exome, further demonstrating their independent clonal origin. We also used *in situ* mutation detection to demonstrate that distinct *KRAS* mutations within a single IPMN regions are spatially separate. Interestingly, a previous study by The Cancer Genome Atlas used computational analyses to suggest that multiple *KRAS* mutations occurred in the same cells in invasive PDACs[94]. The discordant results between these two studies reveal a potential difference between pancreatic precancerous lesions and advanced cancers and highlight the importance of directly measuring individual single-cell genotypes.

Taken together, these data support a revised model of early pancreatic tumorigenesis in which some IPMNs originate from multiple clones that evolve independently, one of which may acquire additional driver gene mutations that lead to clonal expansion and progression (Figure 2.5). Importantly, in the IPMNs originating

from multiple independent clones, our WES data did not identify a single shared mutation preceding *KRAS*, suggesting that there is not an earlier coding genetic event initiating these lesions. Moreover, there were no shared passenger alterations in these distinct sections. Because a subset of such passenger alterations accumulate prior to tumor development, the lack of shared passenger mutations suggests that the different clones arose from completely separate cells[95].

In addition to multiple mutations in early driver genes, we also identified multiple mutations in later-occurring driver genes in the same IPMN. We identified five IPMNs harboring multiple *RNF43* mutations, which were often located in different regions. We also identified one IPMN containing four different *TP53* mutations. Nearly all of the IPMNs with multiple mutations in *TP53* and/or *RNF43* had high-grade dysplasia, suggesting heterogeneity with respect to later-occurring driver gene mutations may be a feature of IPMNs with greater malignant potential. These data add to our revised tumorigenesis model – following the selective sweep and fixation of a dominant clone, additional mutations arise in later-occurring drivers, often multiple mutations in the same gene (Figure 2.5). Similar evidence for convergent evolution has been reported in other tumor types, including colorectal and clear cell renal carcinoma[96,97]. The selective forces that lead to convergence of multiple mutations in the same gene are not known. Further investigation into the conditions that select for multiple mutations within specific genes may provide novel insights into later stages of IPMN progression, as well as the functional outcomes of specific driver gene alterations.

Many studies have demonstrated the value of analyzing cyst fluid collected by endoscopic ultrasound fine needle aspiration (EUS-FNA) from patients diagnosed with

pancreatic cysts. Several reports highlight the diagnostic value of molecular markers in cyst fluid for differentiating likely benign cysts from those which have greater risk of malignant progression[62,63,98]. Integration of our multi-region sequencing data with cyst fluid molecular analysis on a subset of patients in our cohort reveals important clinical implications for interpreting such molecular biomarkers. Not all driver gene mutations within an IPMN may be captured by cyst fluid analysis, even with high-depth sequencing strategies. Because most of the mutations not identified in cyst fluid were present focally in one or a few regions of the IPMN, it is likely that the genetic material containing these mutations was not present in the cyst fluid. These findings highlight the potential challenges with IPMN risk stratification.

There are some limitations to our study. First, while our study encompassed a large cohort of 227 neoplastic tissue samples, these represent only 20 IPMNs. In this study, we prioritized comprehensive analysis of the entire IPMN but could only perform such in-depth analysis on a limited number of IPMN lesions that were entirely submitted for histologic examination. As such, our findings will need to be confirmed in larger IPMN cohorts. Nevertheless, this case size was still sufficient to detect statistically significant differences in genetic heterogeneity between IPMNs with different grades of dysplasia. Second, the majority of samples were sequenced using a small targeted panel of 15 pancreatic driver genes. The use of such a focused sequencing approach allowed us to analyze a large number of samples at high sequencing depth, and thus confidently report the most biologically important driver gene mutations even when they were subclonal. Our sequencing strategy was not designed to provide sufficient data to call copy number alterations or allelic imbalance, and such limitations prevented us from

definitively resolving the evolutionary history of every IPMN. Heterogeneity and evolutionary insights that integrate these types of genetic alterations will require additional studies. Third, our cohort of low-grade and high-grade IPMNs does not have a balanced distribution of histological subtypes, with the gastric histologic subtype enriched in low-grade IPMNs and intestinal and pancreatobiliary subtypes enriched in high-grade IPMNs. However, histologic subtype is not independent of grade of dysplasia, with the vast majority of surgically resected low-grade IPMNs having gastric-type epithelium[99]. Thus, the distribution of histological subtypes in our cohort is representative of low-grade and high-grade IPMNs. Larger studies of genetic heterogeneity specifically focused on histologic subtype may reveal additional insights. Finally, our experimental design required the analysis of surgically resected IPMNs. Because only a small fraction of IPMNs identified radiologically are eventually resected, and the decision to resect is based on specific clinical and radiological criteria[100], our cohort is not representative of the entire spectrum of IPMNs in the population. It is possible that enrichment for larger IPMNs in our resected cohort may also enrich for polyclonality. Analysis of low-grade IPMNs that were not targeted for surgical resection (either through autopsy studies or IPMNs resected along with other neoplasms) will be required to investigate this.

In conclusion, using IPMNs as a system in which to analyze early pancreatic neoplasia, our data provide several insights into the acquisition of somatic mutations during pancreatic tumorigenesis. We show that genetic heterogeneity with respect to driver gene alterations is pervasive in IPMNs. Heterogeneity of the early drivers *KRAS* and *GNAS* is more prevalent during early stages of tumorigenesis, with discordant

KRAS mutations marking multiple independent clones arising from separate cells. In addition, convergent evolution with respect to later-occurring driver gene mutations is present in high-grade IPMNs, demonstrating unique selective pressures at different stages of tumor progression. The results of this study challenge the traditional monoclonal origin of pancreatic tumors, highlighting distinct evolutionary features of precancerous lesions and transforming our understanding of the clonal evolution of pancreatic neoplasia. Future studies will focus on other types of precancerous lesions in the pancreas to determine the broader applicability of these conclusions to pancreatic tumorigenesis.

Table 2.1. Clinicopathological and molecular data from analyzed IPMNs

Case	Sex	Age (at surgery)*	Location	Size (cm)	Predominant Histological Subtype	Final Diagnosis	Tissue Analyzed	# of regions	# of samples (LGD,HGD,PDAC)		Mutations Identified
IPC01	F	80	Head	1.6	Gastric	IPMN with LGD	IPMN	6	6	(6,0,0)	KRAS(p.G12D, p.G12R, p.G12V, p.GQ60GK), RNF43(p.A169T)
IPC02	M	50	Tail	4.5	Gastric	IPMN with LGD	IPMN	8	8	(8,0,0)	KRAS(p.G12D, p.G12R, p.G12V), GNAS(p.R201C, p.R201H), RNF43(p.R286W), TP53(p.P72R)
IPC03	F	60	Tail	2.3	Gastric	IPMN with LGD	IPMN	6	6	(6,0,0)	KRAS(p.G12D, p.G12R, p.G12V), GNAS(p.R201H), CDKN2A(p.R80P)
IPC04	M	70	Head	4.2	Gastric	IPMN with LGD	IPMN	6	6	(6,0,0)	KRAS(p.G12R), GNAS(p.R201H)
IPC05	F	80	Tail	2.8	Gastric	IPMN with LGD	IPMN	8	8	(8,0,0)	KRAS(p.G12V), GNAS(p.R201C)
IPC06	M	60	Body	3	Gastric	IPMN with LGD	IPMN	9	9	(9,0,0)	KRAS(p.G12V), GNAS(p.R201C), BRAF(p.N486_P490del), CTNNB1(p.S45F), TP53(p.L369M)
IPC07	M	60	Head	5	Gastric	IPMN with LGD	IPMN	11	11	(11,0,0)	KRAS(p.G12D, p.G12V), GNAS(p.R201C), RNF43(p.W15*, p.A115Pfs*43), PIK3CA(p.N1044K)
IPC08	F	70	Head	3	Gastric	IPMN with LGD	IPMN	8	8	(8,0,0)	KRAS(p.G12D, p.G12R, p.G12V, p.A146V), GNAS(p.R201C, p.Q227H), RNF43(p.P231L), BRAF(p.N486_P490del), ATM(p.P1526P)
IPC09	M	80	Tail	3	Intestinal	IPMN with LGD	IPMN	7	7	(7,0,0)	KRAS(p.G12D, p.G12V), PIK3CA(p.H1047R)
IPC10	M	70	Head	3	Gastric	IPMN with LGD	IPMN	6	6	(6,0,0)	KRAS(p.G12D, p.G12V), GNAS(p.R201C, p.R201H), RNF43(p.L418M), STK11(p.P203S)
IPC11	M	70	Head	5	Pancreatobiliary	IPMN with HGD	IPMN	10	13	(6,7,0)	KRAS(p.G12V, p.G12S), GNAS(p.R201C, p.R201H, p.Q227E)
IPC12	F	90	Head	5.5	Intestinal	IPMN with HGD	IPMN	23	25	(3,22,0)	KRAS(p.A146V, p.Q61H), GNAS(p.R201C, p.Q227H), RNF43(p.Q6Rfs*9), ATM(p.L2475Yfs*2), TP53(p.P72R)
IPC13	M	70	Head	6.5	Intestinal	IPMN with invasive carcinoma (ductal)	IPMN and carcinoma	22	25	(0,20,5)	KRAS(p.G12D), GNAS(p.R201H)
IPC14	M	70	Head	2.9	Intestinal	IPMN with invasive carcinoma (ductal)	IPMN	11	12	(5,7,0)	KRAS(p.G12R), GNAS(p.R201H), BRAF(p.E26D), ATM(p.L1517F)
IPC15	M	70	Head	3	Pancreatobiliary	IPMN with HGD	IPMN	8	9	(5,4,0)	KRAS(p.G12D, p.G12V), GNAS(p.R201C, p.R201S), RNF43(p.E37*)
IPC16	F	70	Tail	2.4	Intestinal	IPMN with HGD	IPMN	6	10	(7,3,0)	KRAS(p.G12D), GNAS(p.R201H), RNF43(p.R49Sfs*3, p.R132*, p.Q152*, p.T76Nfs*13)
IPC17	M	80	Head	5	Intestinal	IPMN with HGD	IPMN	18	19	(15,4,0)	KRAS(p.A146T, p.Q61H), GNAS(p.R201C), RNF43(p.C119Lfs*6), BRAF(p.T241M), CDKN2A(p.H83N)
IPC18	M	60	Head	4	Pancreatobiliary	IPMN with HGD	IPMN	8	9	(1,8,0)	KRAS(p.G12V), RNF43(p.A115Pfs*43, p.S216L)
IPC19	M	70	Head	4	Intestinal	IPMN with HGD	IPMN	6	6	(0,6,0)	KRAS(p.G12D), GNAS(p.R201H), RNF43(p.A169T, p.R337*), APC(p.R99W)
IPC20	M	70	Head	2.7	Intestinal	IPMN with invasive carcinoma (ductal)	IPMN	24	24	(0,24,0)	KRAS(p.G13D), GNAS(p.R201C), RNF43(p.I141N, p.I201N), TP53(p.V172D, p.P152L, p.D208A, p.R248Q)

*To ensure patient privacy, age (at surgery) was rounded in the nearest decade
LGD: low-grade dysplasia; HGD: high-grade dysplasia

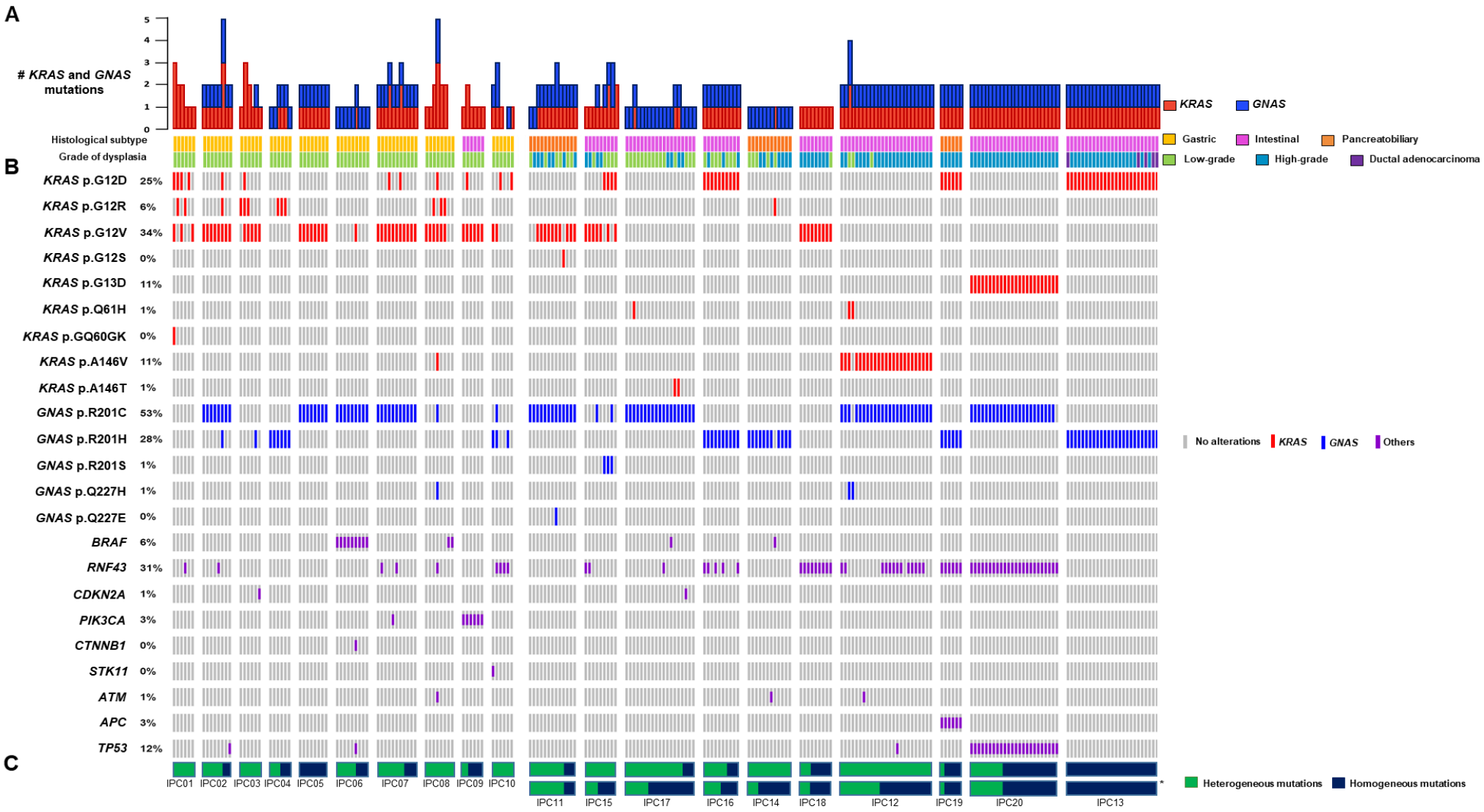

Figure 2.1. Landscape of somatic driver gene mutations in IPMNs. Integrated genomic data for 227 samples (bars) from 20 IPMNs (columns). The histologic subtype and grade of dysplasia data for each sample are shown as tracks at the top. IPMNs are grouped by grade of dysplasia with low-grade on the left and high-grade on the right. **A.** Bar plot representing the total number of *KRAS* and *GNAS* mutations per sample. **B.** Oncoprint heatmap depicting mutations occurring in *KRAS*, *GNAS*, or others (see color legend) or absence (gray bar). Single-nucleotide variants are listed for *KRAS* and *GNAS*, all other mutations are grouped by gene (rows). The percentage of samples with a given mutation is noted at the left. **C.** Proportion of heterogeneous mutations (green) and homogeneous mutations (blue) in each IPMN. *For mixed-grade IPMNs, the proportion of heterogeneous and homogenous mutations was recalculated after excluding low-grade samples (high-grade only)

Table 2.2. Average Jaccard Similarity Coefficient for Each IPMN Based on *KRAS* and *GNAS* Mutations

	Average Jaccard Similarity Coefficient	
Case	Whole Lesion Analysis	High-Grade Only Analysis
IPC01	0.26	
IPC02	0.85	
IPC03	0.38	
IPC04	0.7	
IPC05	1	
IPC06	0.89	
IPC07	0.9	
IPC08	0.28	
IPC09	0.86	
IPC10	0.19	
IPC11	0.74	0.71
IPC12	0.89	1
IPC13	1	1
IPC14	0.83	1
IPC15	0.42	0.75
IPC16	1	1
IPC17	0.85	0.75
IPC18	1	1
IPC19	1	1
IPC20	1	1

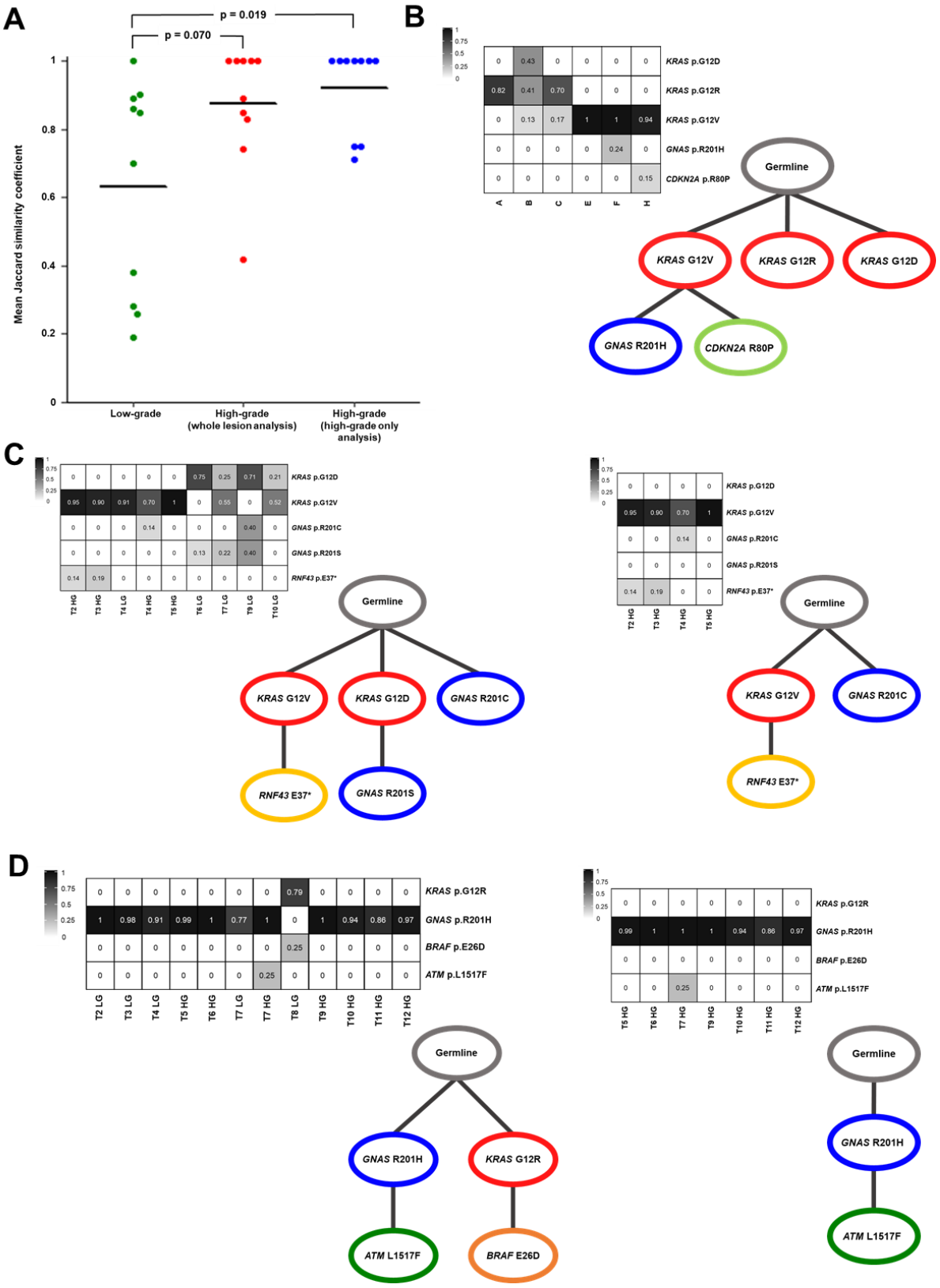

Figure 2.2. Early driver gene heterogeneity and polyclonal origin of IPMNs. A. Dot plot of the average Jaccard similarity coefficient for each case based on *KRAS* and *GNAS* mutations only. Low-grade IPMNs are significantly more heterogeneous than high-grade IPMNs via whole lesion analysis (p value = 0.070; Mann-Whitney *U* test) and high-grade IPMNs via high-grade only analysis (p value = 0.019; Mann-Whitney *U* test). **B-D.** Neoplastic cancer cell fractions (NCFs) are presented in heat maps with each row representing a mutation and each column representing a region or sample. Each NCF heat map corresponds to a schematic of tumor evolution – SCHISM was used to reconstruct somatic mutation hierarchy trees. **B.** Low-grade IPMN, IPC03. **C.** Whole lesion analysis (left) or high-grade only analysis (right) of IPC15. **D.** Whole lesion analysis (left) or high-grade only analysis (right) of IPC14.

Table 2.3. Summary of Somatic Mutations Identified by Whole Exome Sequencing in IPC01

Variant Info						IPC01 DUOD		IPC01 T6		IPC01 T7	
Gene	Variant	Chromosome	Position	Reference allele	Alternate Allele	Distinct Coverage	VAF	Distinct Coverage	VAF	Distinct Coverage	VAF
PRKACA	p.D76N	19	14217583	C	T	39	0.03	119	0.21	209	0.00
CAD	p.A298S	2	27446513	G	T	16	0.00	49	0.12	102	0.00
ABCA13	p.R3914I	7	48431604	G	T	50	0.00	48	0.15	121	0.00
FRRS1	p.T348I	1	100185167	G	A	48	0.00	76	0.28	113	0.00
KRAS	p.G12R	12	25398285	C	G	79	0.00	94	0.27	169	0.00
TAF2	p.F979Lfs20*	8	120759116	A	-	125	0.00	145	0.25	149	0.00
FAM161A	p.N201S	2	62067537	T	C	85	0.00	122	0.24	156	0.00
MYO5A	p.R872*	15	52664524	G	A	23	0.00	35	0.23	62	0.00
ASNS	p.E461_P466del	7	97482450	TGGTCGCCAGAGAATCTC	-	75	0.00	81	0.22	61	0.00
BASP1	p.A203Qfs21*	5	17275928	C	-	83	0.00	196	0.21	146	0.00
ZNF354B	p.H400R	5	178310652	A	G	53	0.00	51	0.21	134	0.00
KDM6A	p.A1230D	X	44949128	C	A	104	0.01	115	0.18	213	0.02
ESPN	p.A466E	1	6505928	C	A	13	0.00	65	0.16	130	0.00
TIAM1	p.V875M	21	32559356	C	T	112	0.01	269	0.00	130	0.19
SDK1	p.T574M	7	4011104	C	T	13	0.00	43	0.00	24	0.17
CENPE	p.T955A	4	104079782	T	C	41	0.00	86	0.00	23	0.52
CASKIN2	p.F982L	17	73498209	G	T	13	0.00	126	0.00	53	0.43
AMPH	p.D655V	7	38429421	T	A	96	0.00	185	0.00	96	0.35
FGF23	p.R196Q	12	4479678	C	T	22	0.00	75	0.00	37	0.25
ZNF831	p.P489L	20	57767540	C	T	45	0.00	285	0.00	125	0.23
ZMYND8	p.W1187G	20	45839414	A	C	53	0.00	218	0.00	93	0.20
TXNRD1	p.T429N	12	104728100	C	A	18	0.00	34	0.00	22	0.18
GATAD2A	p.R324C	19	19606981	C	T	45	0.00	185	0.00	100	0.17
ZNF117	p.H99R	7	64439653	T	C	19	0.00	33	0.00	30	0.17
NFKBIB	p.G279C	19	39398165	G	T	16	0.00	111	0.00	40	0.13
WBSCR17	p.R470L	7	71142200	G	T	15	0.00	68	0.00	32	0.13
A2ML1	p.T947K	12	9008180	C	A	31	0.00	100	0.00	49	0.13
MSH5	p.A349V	6	31725973	C	T	24	0.00	117	0.00	49	0.11
PCDHGA1	p.E561*	5	140711932	G	T	28	0.00	92	0.00	42	0.10
SBNO1	p.N1100K	12	123794396	G	T	19	0.00	62	0.00	42	0.10

Table 2.4. Summary of Somatic Mutations Identified by Whole Exome Sequencing in IPC08

Variant Info						IPC08 DUOD		IPC08 T1		IPC08 T8	
Gene	Variant	Chromosome	Position	Reference allele	Alternate Allele	Distinct Coverage	VAF	Distinct Coverage	VAF	Distinct Coverage	VAF
ZFP36L2	p.F214L	2	43452301	G	T	23	0.04	83	0.35	177	0.00
OPHN1	p.H513Q	X	67316859	G	T	10	0.00	24	0.17	50	0.00
SPTBN2	p.L1779I	11	66458985	G	T	10	0.00	35	0.11	94	0.00
TTC8	p.E344*	14	89337993	G	T	24	0.00	52	0.37	103	0.00
CSMD3	p.R1041K	8	113662461	C	T	34	0.00	77	0.36	106	0.00
TRIM41	p.R450C	5	180661230	C	T	20	0.00	47	0.17	116	0.00
KCNN1	p.V273F	19	18092836	G	T	11	0.00	39	0.10	140	0.00
NEB	p.N6001Y	2	152423837	T	A	55	0.00	97	0.50	186	0.00
FASTKD5	p.A98T	20	3129425	C	T	35	0.00	74	0.47	167	0.00
ZFHX4	p.S565R	8	77618016	A	C	14	0.00	48	0.46	97	0.00
RGAG4	p.E516Q	X	71349845	C	G	26	0.00	62	0.42	169	0.00
KRAS	p.G12V	12	25398284	C	A	48	0.00	58	0.35	184	0.00
ATR	p.P2188S	3	142186901	G	A	17	0.00	16	0.25	59	0.00
LRRC3C	p.S75P	17	38100382	T	C	11	0.00	30	0.21	80	0.00
SERPINA7	p.V367F	X	105277640	C	A	14	0.00	26	0.19	48	0.00
C12orf5	p.E19D	12	4440470	G	T	24	0.00	34	0.15	74	0.00
MSH4	p.A437S	1	76342624	G	T	17	0.00	40	0.13	30	0.00
SVEP1	p.A2926S	9	113169104	C	A	20	0.00	42	0.12	82	0.00
KRTAP26-1	p.D28Y	21	31692272	C	A	19	0.00	33	0.12	83	0.00
BEND2	p.R210I	X	18221899	C	A	20	0.00	43	0.12	105	0.00
MYL6B	p.E116D	12	56549204	G	T	24	0.00	41	0.12	94	0.03
PANK4	p.V577F	1	2443121	C	A	23	0.00	45	0.11	87	0.00
SCAPER	p.K985N	15	76763676	C	A	30	0.00	38	0.11	116	0.02
MLLT4	p.S238R	6	168276153	C	A	29	0.00	39	0.10	122	0.00
SRCAP	p.Y621*	16	30723630	C	A	14	0.00	50	0.10	112	0.00
IRS4	p.A223V	X	107978907	G	A	10	0.00	42	0.10	102	0.00
DNAJB13	p.P33Q	11	73669391	C	A	21	0.00	92	0.00	63	0.27
OR1E2	p.M85I	17	3336881	C	A	19	0.00	43	0.00	30	0.17
RIOK3	p.A337E	18	21053587	C	A	13	0.00	41	0.00	30	0.17
C4BPA	p.Y104H	1	207287612	T	C	20	0.00	60	0.00	41	0.15
ATP10D	p.P88Q	4	47514820	C	A	14	0.00	71	0.00	41	0.15
KIAA1549	p.S1801*	7	138529112	G	T	22	0.00	48	0.00	38	0.14
KHDC1	p.A205S	6	73951353	C	A	10	0.00	37	0.00	30	0.13
RRN3	p.S548L	16	15159139	G	A	59	0.00	157	0.00	132	0.13
FLT3	p.T255I	13	28623890	G	A	18	0.00	61	0.00	35	0.11
SRGAP2	p.V150I	1	206566903	G	A	34	0.00	133	0.02	63	0.11
ATR	p.D883Y	3	142272227	C	A	50	0.00	122	0.00	70	0.10

Table 2.5. Summary of Somatic Mutations Identified by Whole Exome Sequencing in IPC12

Variant Info						IPC12 DUOD		IPC12 T3		IPC12 T6	
Gene	Variant	Chromosome	Position	Reference allele	Alternate Allele	Distinct Coverage	VAF	Distinct Coverage	VAF	Distinct Coverage	VAF
ZSWIM6	p.Q7*	5	60628118	C	T	13	0.00	23	0.39	36	0.00
HHIPL2	p.V694I	1	222696038	C	T	64	0.02	89	0.21	101	0.00
ANO2	p.S724*	12	5708727	G	T	51	0.02	60	0.10	64	0.00
PNMA3	p.R389G	X	152226577	A	G	21	0.00	44	0.37	74	0.00
PSD2	p.A55V	5	139189189	C	T	27	0.00	38	0.16	75	0.00
BLM	p.A3V	15	91290630	C	T	40	0.00	54	0.13	97	0.00
GTF3C3	p.R819H	2	197631371	C	T	91	0.01	72	0.13	96	0.00
TAF5	p.S765L	10	105147871	C	T	157	0.01	172	0.37	222	0.00
GNAS	p.R201C	20	57484420	C	T	91	0.00	133	0.63	175	0.02
KRAS	p.A146V	12	25378561	G	A	53	0.00	56	0.61	79	0.00
KIAA1462	p.A990V	10	30316108	G	A	49	0.04	37	0.60	83	0.02
INPPL1	p.L806H	11	71946161	T	A	27	0.00	20	0.55	50	0.00
OBSCN	p.A6815T	1	228528464	G	A	39	0.03	67	0.51	95	0.00
SYNE1	p.V172M	6	152831395	C	T	40	0.00	45	0.49	57	0.00
ESYT3	p.R309S	3	138183198	A	C	28	0.00	41	0.49	53	0.00
ITSN2	p.I1518M	2	24432010	A	C	24	0.00	29	0.48	43	0.00
CDH1	p.I326L	16	68845730	A	C	38	0.00	41	0.46	73	0.00
TBATA	p.R54Q	10	72541673	C	T	41	0.05	57	0.44	118	0.01
FOXA3	p.F171S	19	46375775	T	C	26	0.00	66	0.42	72	0.00
HSCB	p.S117P	22	29140618	T	C	87	0.00	87	0.42	142	0.00
RNF43	p.Q6Rfs9*	17	56492913	AGCTGCAGCT	-	117	0.00	70	0.40	58	0.00
RHOBTB3	.	5	95124448	TTTTCATTTTTCTTGCAGT	-	330	0.00	268	0.39	222	0.00
GRIN2A	p.V617M	16	9923438	C	T	75	0.00	75	0.37	90	0.00
KIRREL	p.A166T	1	158054355	G	A	37	0.03	43	0.37	61	0.00
PLEKHG3	p.E443D	14	65205552	G	T	16	0.00	46	0.36	71	0.00
PAX3	p.A410V	2	223066851	G	A	51	0.02	44	0.34	36	0.00
IGKV3-11	p.I78V	2	89326781	T	C	26	0.00	22	0.33	28	0.00
KIAA0947	p.L1752R	5	5464702	T	G	54	0.00	54	0.32	79	0.00
HEPH	p.C801*	X	65423369	C	A	58	0.00	55	0.29	47	0.00
SLC22A23	p.E590A	6	3273581	T	G	43	0.00	86	0.25	87	0.00
LPIN3	p.I438V	20	39980809	A	G	63	0.00	140	0.22	175	0.00
ARMC3	.	10	23257419	G	A	28	0.04	19	0.21	12	0.00
DENND4C	p.I1595Hfs3*	9	19357115	-	T	282	0.00	349	0.21	259	0.00
GLYATL2	p.F256C	11	58602020	A	C	33	0.00	25	0.20	43	0.00
DCTN2	p.M1?	12	57940846	T	A	15	0.00	23	0.17	17	0.00
C3orf84	p.D59N	3	49215938	C	T	19	0.00	34	0.12	31	0.00
RNFT2	p.R416H	12	117287165	G	A	32	0.00	68	0.12	61	0.00
MAGEA1	p.A162V	X	152482526	G	A	80	0.01	108	0.11	174	0.00
SPTAN1	p.R2303H	9	131394566	G	A	15	0.00	41	0.10	65	0.00
GNAS	p.Q227H	20	57484597	G	C	50	0.00	138	0.00	69	0.26
ITGA2	p.D501Y	5	52358658	G	T	170	0.01	301	0.00	160	0.20
KRAS	p.Q61H	12	25380275	T	G	60	0.00	123	0.00	50	0.26
ZW10	p.L375V	11	113618402	G	C	66	0.00	95	0.00	46	0.24
RASSF6	p.L60V	4	74464419	A	C	52	0.00	107	0.00	47	0.22
JAK2	p.W157*	9	5050687	G	A	142	0.00	268	0.00	100	0.19
ACVR1B	p.N381T	12	52379015	A	C	41	0.00	131	0.00	44	0.18
RND2	p.R54C	17	41178034	C	T	16	0.00	81	0.00	30	0.17
AKT3	p.E84K	1	243828108	C	T	31	0.00	103	0.00	52	0.15
ZNF573	p.T193K	19	38230813	G	T	34	0.00	52	0.00	43	0.14
OPLAH	p.K92N	8	145114589	C	A	18	0.00	63	0.00	46	0.11
BDH1	p.N168Y	3	197241195	T	A	99	0.00	284	0.00	81	0.10

Table 2.6. Summary of Somatic Mutations Identified by Whole Exome Sequencing in IPC14

Variant Info						IPC14 DUOD		IPC14 T5		IPC14 T8	
Gene	Variant	Chromosome	Position	Reference allele	Alternate Allele	Distinct Coverage	VAF	Distinct Coverage	VAF	Distinct Coverage	VAF
CEACAM16	p.A196S	19	45207491	G	T	14	0.00	34	0.21	51	0.00
GNAS	p.R201H	20	57484421	G	A	70	0.00	154	0.50	190	0.00
HCFC1	p.E1783K	X	153217072	C	T	34	0.00	64	0.11	99	0.00
FAM189B	p.V157I	1	155223694	C	T	49	0.00	121	0.20	208	0.00
HIST1H1B	p.A179T	6	27834773	C	T	109	0.01	131	0.15	187	0.00
ZBED2	p.R4Q	3	111313038	C	T	68	0.02	113	0.10	214	0.00
ACACB	p.L1398Pfs26*	12	109671606	T	-	246	0.00	293	0.60	364	0.00
UBE2NL	p.M1N	X	142967203	A	G	19	0.00	36	0.78	31	0.00
ZBTB21	p.P436L	21	43412898	G	A	37	0.00	41	0.60	93	0.00
SMARCB1	p.T357Dfs4*	22	24175838	-	T	42	0.00	79	0.48	135	0.00
ESRP2	p.P125A	16	68267965	G	C	32	0.00	86	0.46	106	0.00
HTR3C	p.Q37H	3	183772552	G	C	53	0.00	73	0.39	69	0.00
POLR2B	p.R1141H	4	57896552	G	A	45	0.00	44	0.39	87	0.00
CCDC36	p.K176T	3	49282161	A	C	39	0.00	86	0.36	132	0.00
GRID1	p.A738T	10	87379772	C	T	44	0.00	44	0.33	102	0.00
WDFY4	p.L2168H	10	50038907	T	A	19	0.00	28	0.25	54	0.00
GNL2	p.E644G	1	38033896	T	C	44	0.00	81	0.25	81	0.00
PCSK6	p.R225*	15	101970251	G	A	40	0.03	41	0.24	104	0.00
RICTOR	p.I920M	5	38953593	A	C	41	0.00	88	0.23	149	0.00
TARDBP	p.C39Y	1	11073900	G	A	51	0.00	73	0.22	86	0.00
TOPBP1	p.H735D	3	133358833	G	C	118	0.00	179	0.22	258	0.00
ADAMTS9	p.C553Y	3	64633668	C	T	35	0.00	56	0.21	63	0.00
ENTPD8	p.L479V	9	140329419	G	C	45	0.00	144	0.19	248	0.00
NABP2	p.T40S	12	56619195	A	T	75	0.03	176	0.18	191	0.00
KRT2	p.R122H	12	53045562	C	T	45	0.00	89	0.18	102	0.00
OR5K3	p.F32Lfs57*	3	98109599	-	T	72	0.00	84	0.18	75	0.00
CLDN18	p.T4Sfs140*	3	137717718	TGACTGCC	-	63	0.00	108	0.18	57	0.00
SRD5A2	p.S41R	2	31805847	G	T	14	0.00	24	0.17	49	0.00
FLRT2	p.A503V	14	86089366	C	T	65	0.02	88	0.15	164	0.00
PIP4K2C	p.Q36H	12	57985180	G	C	56	0.00	241	0.15	249	0.00
PEX19	p.N284K	1	160249389	A	C	28	0.00	35	0.14	46	0.00
KAL1	p.Q196H	X	8555973	C	A	37	0.03	49	0.14	56	0.00
TTC21B	p.T1165M	2	166740494	G	A	108	0.02	127	0.13	173	0.00
ARRB1	p.L339M	11	74979987	G	T	43	0.02	30	0.13	117	0.00
CAPRIN2	p.S278*	12	30886622	G	C	129	0.00	397	0.13	216	0.00
VPS33B	p.R290W	15	91549274	G	A	27	0.00	32	0.13	48	0.00
RAPSN	p.A265D	11	47463281	G	T	33	0.00	80	0.13	121	0.00
UBE4A	p.L813W	11	118257157	T	G	32	0.00	41	0.12	90	0.00
BARD1	p.W629*	2	215609807	C	T	49	0.00	58	0.12	115	0.00
OPN3	p.D191H	1	241767684	C	G	65	0.00	105	0.11	129	0.00
KRT2	p.R275C	12	53043736	G	A	117	0.01	146	0.10	218	0.00
BTNL9	p.D109N	5	180475142	G	A	41	0.00	61	0.10	101	0.00
COBLL1	p.S610C	2	165555957	T	A	229	0.00	262	0.10	461	0.00
BCL11A	p.I832M	2	60687551	T	C	48	0.00	91	0.00	57	0.30
SLC40A1	p.S301P	2	190428811	A	G	84	0.00	146	0.00	65	0.23
VCAN	p.S1634T	5	82833722	T	A	60	0.02	135	0.00	63	0.22
WNT16	p.G157S	7	120971854	G	A	106	0.00	268	0.00	126	0.19
GPR98	p.S3562N	5	90040998	G	A	62	0.00	91	0.00	38	0.13
SPOCD1	p.G1136C	1	32256449	C	A	19	0.00	64	0.00	40	0.13
PRPF8	p.L380I	17	1583054	G	T	23	0.00	69	0.00	44	0.11
FBP1	p.R255H	9	97367800	C	T	36	0.00	86	0.00	46	0.11
APLF	p.H307N	2	68765118	C	A	54	0.00	132	0.00	68	0.10
ACBD5	p.W511C	10	27493401	C	A	88	0.00	233	0.00	107	0.33
LARP6	p.K302*	15	71124963	T	A	76	0.00	192	0.00	151	0.27
SCAPER	p.T451S	15	77059327	T	A	131	0.00	271	0.00	137	0.22
KRAS	p.G12R	12	25398285	C	G	101	0.00	589	0.00	110	0.20

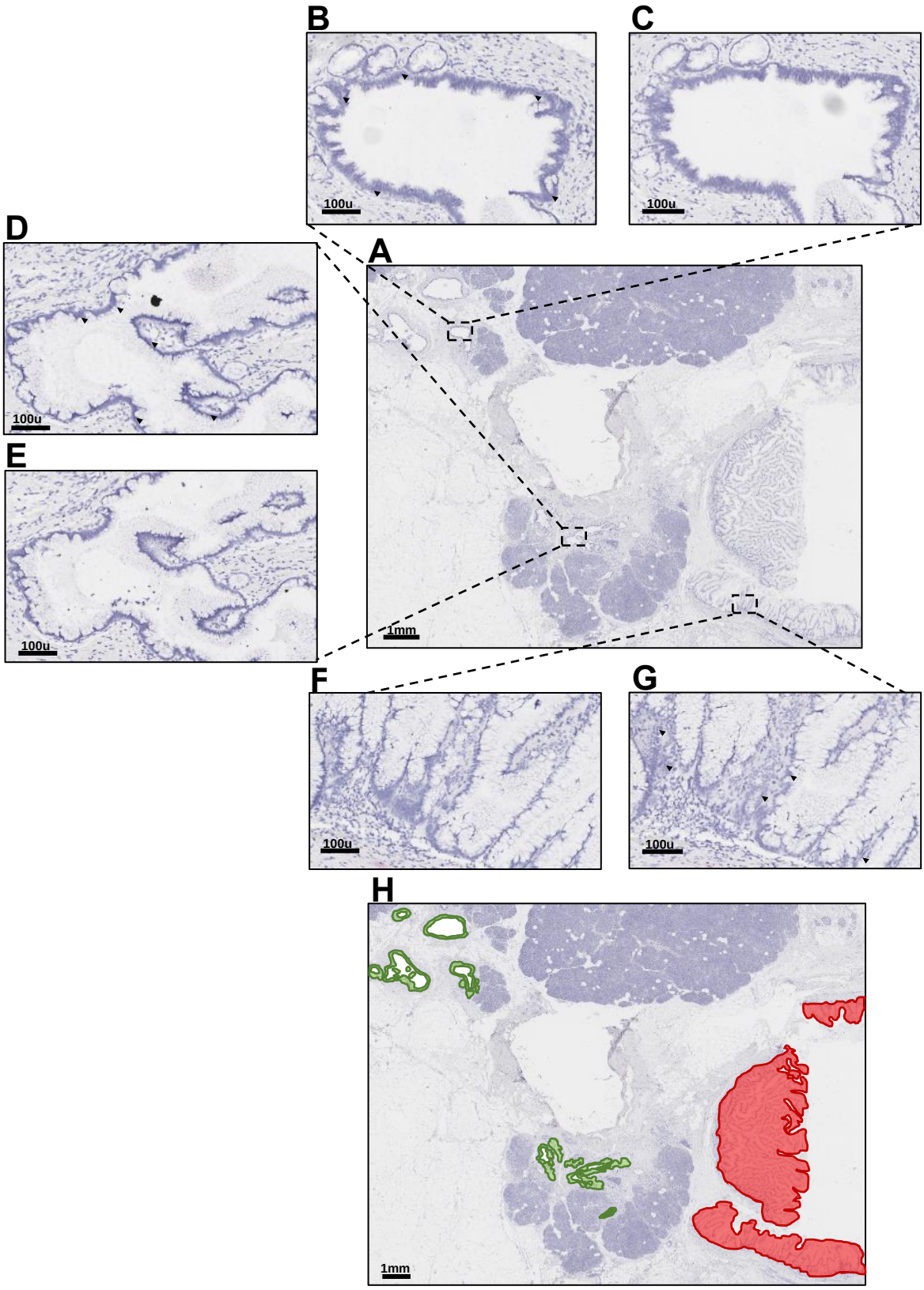

Figure 2.3. BaseScope *in-situ* RNA detection of mutant *KRAS* in FFPE IPMN tissues. A. Hematoxylin stain of low-grade IPMN sample IPC09 B. **B-C.** Representative image of Area 1 probed for either *KRAS* p.G12D (**B**) or *KRAS* p.G12V (**C**). **D-E.** Representative image of Area 2 probed for either *KRAS* p.G12D (**D**) or *KRAS* p.G12V (**E**). **F-G.** Representative image of Area 3 probed for either *KRAS* p.G12D (**F**) or *KRAS* p.G12V (**G**). **H.** Hematoxylin stain of IPC09 B overlaid with green, indicating areas of positive staining for *KRAS* p.G12D or red, indicating areas of positive staining for *KRAS* p.G12V. Dark red dots designate positive signal (black arrows point to representative cells with positive staining). Nuclei were counterstained with hematoxylin. **A** and **H** = Magnification 0.5X. **B-G** = Magnification 15X.

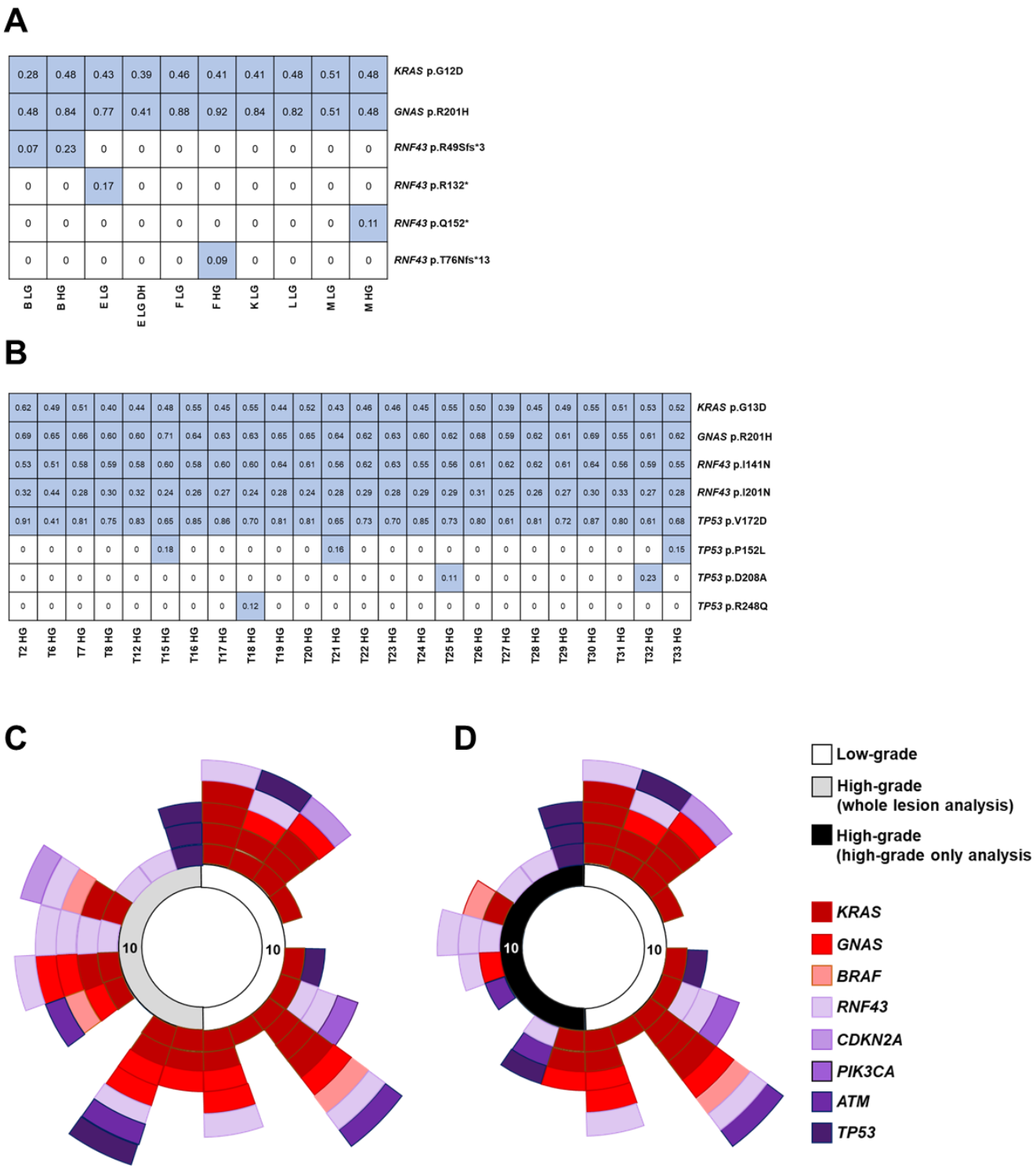

Figure 2.4. Convergent evolution and heterogeneity in later-occurring driver genes in IPMNs with high-grade dysplasia. A-B. Mutations are presented in tables with each row representing a mutation and each column representing a sample. Blue color indicates a mutation call, with variant allele frequencies in each cell of the table. Both depicted IPMNs have multiple mutations called in a later-driver gene. **A.** IPC16, a high-grade IPMN, has four distinct *RNF43* mutations. **B.** IPC20, a high-grade IPMN, has two different *RNF43* mutations and four distinct *TP53* mutations. **C-D**. The 20 IPMNs are represented around the perimeter of the circle in clockwise numerical order – grade of dysplasia/analysis-type for each IPMN is indicated by the colored key. Only heterogeneous mutations are displayed (not all mutations). Each colored rectangle represents a heterogeneous mutation in a given gene, which is indicated by colored key. The heterogeneous mutations in each IPMN are presented by gene in low-grade and whole lesion analysis of high-grade IPMNs (**C**) and low-grade and high-grade only analysis of high-grade IPMNs (**D**). Note: Only genes with mutations in at least two IPMNs were included in C-D.

Table 2.7. Comparison of Mutations Identified in Multi-Region Sequencing and Cyst Fluid Molecular Analysis in IPC03

Mutations called in cyst fluid and multi-region sequencing				Mutations called only in multi-region sequencing				Mutations called only in cyst fluid	
Mutation	VAF cyst fluid	VAF multi-region	# of regions with mutation present at >5% VAF	Mutation	VAF multi-region	# of regions with mutation present at >5% VAF	Region covered by cyst fluid	Mutation	VAF cyst fluid
KRAS p.G12V	0.27	0.05-0.45	5	KRAS p.G12R	0.16-0.33	3	Y		
				KRAS p.G12D	0.17	1	Y		
				GNAS p.R201H	0.10	1	Y		
				CDKN2A p.R80P	0.06	1	N		

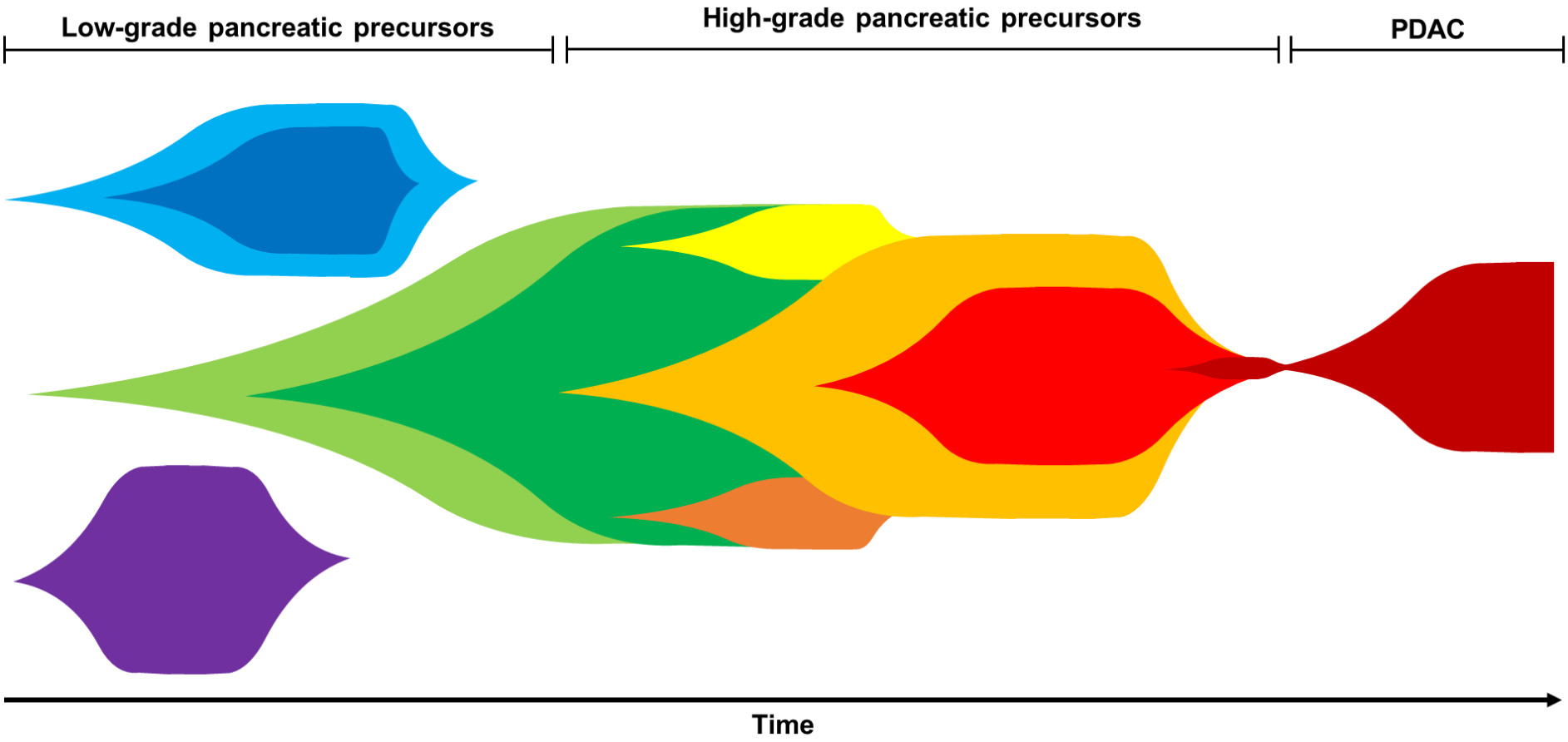

Figure 2.5. Revised clonal evolution model for early pancreatic tumorigenesis. Multiple, independent clones arise with distinct mutations in early driver genes (i.e. *GNAS* and/or *KRAS*) (blue, purple, and light green). Early in tumorigenesis, some selective pressures eventually lead to expansion of a dominant clone (green). Often during later-stages of tumorigenesis, convergent evolution leads to multiple mutations in the same later-occurring driver gene (i.e. *RNF43*, *TP53*) (yellow, gold, and orange). Eventually, a dominant clone invades giving rise to pancreatic ductal adenocarcinoma (dark red).

CHAPTER 3:

Single-cell sequencing reveals genetic heterogeneity of pancreatic precancers

Rationale

Single-cell sequencing can resolve individual cellular genotypes and definitively assign mutations to each clone, providing a more complete understanding of intratumoral genetic heterogeneity. Multiple recent studies have reported reliable single-cell sequencing of human cancer specimens[101–104]. However, to our knowledge, the genotypes of individual cells in pancreatic precursor lesions have yet to be examined. In this study, we assess genetic heterogeneity in IPMNs by identifying somatic mutations in single neoplastic cells from fresh IPMN tissue from surgical resections, providing new insights into the clonal structure of these pancreatic cancer precursor lesions

Materials and Methods

***Specimen acquisition*.** This study was approved by the Institutional Review Board at The Johns Hopkins Hospital. We prospectively collected samples from ten patients diagnosed with IPMN who had undergone surgical resection at the Johns Hopkins Hospital between 2015 and 2016. Neoplastic tissue was harvested by a subspecialized pancreatic pathologist; fragments of cyst call were harvested from one to four different regions of IPMN after confirmation of diagnosis.

Tissue processing and isolation of single cells. The harvested neoplastic tissues were immediately minced using a sterile scalpel and tweezers. Minced tissues were then enzymatically digested using optimized MEM containing 375 units/ml collagenase type 4, 250 units/ml hyaluronidase, 2.4 units/ml dispase II, and 1% FBS, and incubated in a shaking incubator at 37C for 1 hour. Subsequently, the digested cell solution was filtered through a 70um mesh and washed with Hank’s Balanced Salt Solution (HBSS)

with 10% FBS. Fresh normal duodenum or spleen was harvested at the same time, flash frozen, and stored at -80C until DNA extraction.

Fluorescence-activated cell sorting. The single cell suspension was adjusted to the concentration of 1 million cells/200ul HBSS containing 5% FBS. The cells were labeled with 1.5ug/ml Anti-EpCAM antibody [VU-1D9] (Phycoerythrin) (abcam, ab112068) and incubated on ice for 30 min. Following centrifugation, cells were washed twice with HBSS. Cells were resuspended and adjusted to the concentration of 1 million cells/1ml containing 2% FBS, and then labeled with 5uM Vybrant® DyeCycle™ Violet Stain (Life Technologies, V35003) and incubated using a water bath at 37°C for 30 min. Before sorting, 1ug/ml propridium iodide (PI) (Sigma Aldrich, P4864) was added to the cell solution, which was then filtered through 35um mesh. Single viable epithelial cells in G2/M phase were sorted using MoFlo Legacy cell sorter (Beckman Coulter, Inc., Miami, FL) into individual wells of a 96-well plate. First, we selected intact cells based on forward scatter (FS) and side scatter (SS), and we then selected the population to exclude PI-positive cells to ensure the presence of only viable cells. We subsequently selected EpCAM-positive cells. In addition, as first reported by Wang and colleagues[103], we specifically selected cells that had replicated their DNA to improve the efficiency of the subsequent whole genome amplification – this was accomplished with DyeCycle Violet staining and selection of cells in G2/M phase, according to Vybrant® DyeCycle™ Violet Stain intensity (using the EpCAM-negative population as a control for G1 DNA content). We isolated 23-46 single cells per harvested tissue section, for a total of 23-92 single cells per patient. One bulk sample of 1,000 cells was

isolated for each tissue sample, and when an adequate number of cells were available, a second bulk sample of 30,000 cells was also isolated for each tissue sample.

Preparation of genomic DNA. Multiple displacement amplification (MDA) using isothermal random priming and extension with φ29 polymerase was conducted on the sorted single cells using the REPLI-g Single Cell kit (Qiagen, 150345) according to the manufacturer's instructions - this WGA approach leads to higher coverage and lower error rate compared to other amplification approaches[105,106]. The same kit was used to amplify DNA from the 1,000-cell bulk sample. Amplified DNA was purified by the ethanol precipitation method according to the manufacturer's protocol. Genomic DNA from the sorted bulk cells was isolated using the QIAamp® DNA Micro Kit (Qiagen, 56304), and genomic DNA from frozen normal tissue was isolated using the QIAamp® DNA Mini Kit (Qiagen, 51304) according to the manufacturer's instructions. All prepared samples were quantified by quantitative PCR (qPCR) targeting long interspersed nuclear elements (LINE) using StepOnePlus™ Real-Time PCR System (ThermoFisher scientific) {Rago, 2007 #395}. PCR was performed using Power SYBR Green Master Mix (ThermoFisher scientific, 4367659) according to the manufacturer's instructions. The LINE primes were designed as follows: 5'-AAAGCCGCTCAACTACATGG-3' (forward) and 5'-TGCTTTGAATGCGTCCCAGAG-3' (reverse). The real-time PCR conditions were as follows: 50°C for 2 minutes, 94°C for 2 minutes, 40 cycles of 94°C for 10 seconds, 58°C for 15 seconds, and 70°C for 30 seconds, 95°C for 15 seconds, 60°C for 30 seconds, and 95°C for 15 seconds.

Evaluation of amplification efficiency. To determine quality of the WGA products for downstream experiments, multiplex PCR was conducted using the Multiplex PCR Plus

Kit (Qiagen, 206152). Ten pairs of primers were designed to target ten genes located on different chromosomes. The PCR conditions were as follows: 95°C for 5 minutes, 35 cycles of 95°C for 30 seconds, 63°C for 3 minutes, and 72°C for 40 seconds, and 68°C for 10 minutes. PCR products were run on 2% agarose gels. WGA products were selected as successfully amplified products by detection of the DNA bands in at least 8/10 (80%) PCR products[104,107,108].

Targeted deep sequencing in 11 genes. Using the Ion AmpliSeq Designer, we designed an Ion AmpliSeq custom panel 142 primer pairs in 11 genes reported previously: *ARID1A, TGFBR2, PIK3CA, BRAF, CDKN2A, KRAS, TP53, MAP2K4, RNF43, SMAD4,* and *GNAS*. Additional primers were designed to target single nucleotide polymorphisms (SNPs) with high minor allele frequency archived by National Center for Biotechnology Information (NCBI) in or near these genes, for a total of 166 primer pairs in the panel. Sample DNA (20ng/primer pool) was amplified and the libraries were prepared using Ion AmpliSeq 2.0 Library Kit (Life Technologies) according to the manufacturer's protocol. Individual samples were barcoded, pooled, and templated with the OneTouch2 system (Life technologies), and subsequently sequenced on the Ion Torrent Personal Genome Machine (Life Technologies). Targeted sequencing was performed on all single cells that amplified 8/10 products in our multiplex PCR assay (see above), one bulk tumor sample per tissue section, and one bulk normal sample per patient. For IP10 section B, IP10 section C, and IP16 section C, we used an unamplified tumor bulk sample (from 30,000 sorted cells). For the remaining samples, we used an amplified tumor bulk sample (from 1,000 sorted cells followed by

WGA as described above) because we were unable to sort an adequate cell number for an unamplified bulk sample.

Mutation calling and imputation of undetermined genotypes. For each IPMN single-cell, bulk tumor, or bulk normal sample, reads were mapped to the human reference genome version hg19 with BWA-mem[81], and the resulting BAM files were processed using the Genome Analysis ToolKit (GATK). For each of the single cell BAMs, variant calling was done with GATK HaplotypeCaller (independently for each one)[80]. Each variant site was classified as either harboring a somatic mutation (VAF≥5% and alternate read count ≥ 5, and normal bulk VAF<2% and alternate read count <4, provided that normal bulk coverage ≥100X), consistent with reference (coverage≥100 and VAF<0.2% and alternate read count <2), or indeterminate. Mutation profile matrices (mutation by single-cell) were constructed – values of 0, 1, or 0.5 were assigned to each matrix element, depending on whether the mutation was present, absent, or if its status was indeterminate, in the corresponding single cell. Only mutations called in at least two single cells or one single cell and one bulk IPMN sample were considered in further analysis. To impute indeterminate elements, we applied an iterative single-cell genotype clustering algorithm using minimum distance linkage to impute the status of indeterminate mutations[109].

Allelic dropout estimation*.* The normal bulk samples were used to estimate single-cell allelic dropout (ADO) rates. As described above, we applied GATK's HaplotypeCaller to call sites in the normal bulk harboring a germline variant and mpileup to compute VAFs and alternate read counts[110]. Each site called by HaplotypeCaller as heterozygous in the normal bulk was examined to see if it was homozygous in each single cell (VAF<5%

or VAF>95%). A single cell's ADO rate was computed as the fraction of heterozygous sites in the bulk normal that had sufficiently high coverage (≥100x) but were found to be homozygous in that single cell. We removed single cells with fewer than 51% heterozygous SNPs correctly called as heterozygous. The 51% threshold was determined by standard box-plot techniques. Finally, single cells for which fewer than 5 of the heterozygous sites in the bulk normal had sufficiently high coverage were also excluded from further analysis, as their ADO rate could not be accurately determined. The average allelic dropout (ADO) rate across all single cells was 13% after outlier removal.

***Analysis of technical controls*.** Three different technical controls were processed as described above in order to calculate the false positive rate (FPR) of our single-cell analysis pipeline. First, 28 single cells from the Pa01 cell line were sorted as described above, followed by whole genome amplification and multiplex PCR to assay amplification efficiency. In addition, four bulk samples of this cell line were analyzed, including one sample of 5×10^6 cells without amplification and three samples of 10 cells that underwent whole genome amplification. Single-cell and bulk samples were sequenced using the Ion AmpliSeq Cancer Hotspot Panel v2, and mutations were called as described above. In the absence of a matched normal sample, we could not distinguish somatic mutations. Instead, the mutations called in the bulk samples (which included both germline and somatic mutations) were considered true positives, while any mutation called in a single cell but absent in the bulk was considered a false positive – this resulted in a FPR of 2.5×10^{-5} false positives per base pair sequenced. The use of a cell line for technical control has the advantage of genetic homogeneity;

however, it does not account for potential errors induced by tissue processing steps. As such, we included two additional technical controls derived from resected human pancreatic samples. First, we used normal cells isolated from PDAC samples. Specifically, tissue was harvested from pancreata from 6 PDAC patients, followed by tissue digestion, single cell isolation, whole genome amplification, multiplex PCR, and targeted sequencing using our AmpliSeq custom panel as described above. Mutations were called as described above. Due to the low neoplastic cellularity of PDACs, a sizeable proportion of the analyzed cells were non-neoplastic. Thus, we defined "normal" cells as those that had no somatic mutations that were present in the bulk sample from the same patient. Any mutation called in a "normal" cell was considered a false positive – in total, we analyzed 23 single normal cells from 3 PDAC samples and calculated a FPR of 3.0×10^{-5} false positives per base pair sequenced. As a third technical control, we used single cells sequenced from an intraductal oncocytic papillary neoplasm (IOPN), which were processed and sequenced with an identical protocol to the our experimental IPMNs. Because IOPNs lack mutations in the pancreatic driver genes in our custom panel and all bulk samples from this lesion lacked somatic mutations, we considered any somatic mutation in this lesion to be a false positive – we calculated a FPR of 3.1×10^{-5} false positives per base pair sequenced. Importantly, the FPR is consistent across these three independent technical controls, suggesting that it is protocol-specific and does not vary substantially between different experimental samples. Moreover, the same false positive mutation was not called in multiple single cells in any of the technical controls, highlighting that the mutations identified in our experimental samples are unlikely to be technical artifacts.

Results

Clinicopathological data. Clinicopathological data of the ten cases are summarized in Table 3.1. All patients had a histologically confirmed diagnosis of IPMN. Of the ten IPMNs analyzed, seven cases were non-invasive (three with low-grade dysplasia and four with high-grade dysplasia), while three had adjacent invasive carcinomas (two invasive ductal adenocarcinomas and one invasive colloid carcinoma). The IPMNs were classified as gastric (7 cases), intestinal (2 cases), and oncocytic (1 case) histological subtypes. The number of somatic mutations called in the ten analyzed IPMNs ranged from 0 to 10 (Figures 3.1-3.4). Five IPMNs had different mutations in the same driver gene called in different single cells.

Early driver gene mutations. Of the ten IPMNs analyzed, nine harbored *KRAS* and/or *GNAS* mutations (Table 3.1). We identified two IPMNs with genetic heterogeneity with respect to these early driver gene mutations. One case (IP27) was diagnosed as an IPMN with high-grade dysplasia - two pieces of the cyst lining were harvested from the same grossly-evident cyst. Two different distinct *KRAS* mutations were identified in this neoplasm (Figure 3.1A). The majority of cells that harbored a *GNAS* p.R201H mutation also had a *KRAS* p.G12D mutation. Three cells lacked both the *GNAS* p.R201H and *KRAS* p.G12D mutations but had a different mutation in *KRAS* (p.G12R). These data suggest this neoplasm could have originated from two independent clones, each with unique mutations in early driver genes and without any shared genetic alterations. In the second IPMN (IP16), there were two adjacent grossly-evident cysts (samples A and B) as well as a distinct firm area (sample C). Histologic examination revealed IPMN with high-grade dysplasia that was microscopically contiguous between the two cysts

(suggesting that they likely represented the same IPMN) and infiltrating ductal adenocarcinoma in the firm area. Two different clones with unique *KRAS* mutations were identified in the cysts – a clone with *KRAS* p.G12D was identified in both IPMN samples, while a clone with *KRAS* p.G12V was limited to a single IPMN sample. A mutation in *TP53* (p.R175H) occurred in a subset of cells with *KRAS* p.G12D. In contrast, the invasive cancer harbored a different *KRAS* mutation (p.G12R) as well as a unique *TP53* mutation (p.C229*) (Figure 3.1B). These results suggest two independent clones within the IPMNs, as well as a genetically distinct invasive adenocarcinoma, all in a relatively small area of this patient's pancreas.

In the other seven cases, shared *KRAS* and/or *GNAS* mutations were present in the vast majority of neoplastic cells analyzed, suggesting that these were clonal mutations.

Mutations in other driver genes. We identified mutations in *RNF43* in three IPMNs (Table 3.1). The first *RNF43*-mutant case (IP22) was diagnosed as IPMN with low-grade dysplasia and harbored six different *RNF43* mutations, including three frameshift insertion/deletion mutations, two missense mutations, and one binucleotide substitution. The binucleotide substitution (p.VD299VY) occurred in the same clone as one missense mutation (p.G166V), suggesting biallelic *RNF43* alteration in this clone. The remaining *RNF43* mutations were mutually exclusive, suggesting the presence of five separate clones with unique alterations in *RNF43* (Figure 3.2A). In another IPMN with high-grade dysplasia (IP24), we identified three *RNF43* mutations each present in different single cell populations: p.A11Lfs occurred in three single cells from section A, while p.G207D occurred in a single cell from section B and p.T28I occurred in one single cell from each

tissue section (Figure 3.2B). In the third *RNF43*-mutant neoplasm (IP10), which was diagnosed as colloid carcinoma, the *RNF43* mutation (p.L12Dfs) occurred with mutations in *GNAS* (p.R201C) and *CDKN2A* (p.G55Afs) in all cells analyzed from this IPMN (Figure 3.4). However, we analyzed very few single cells in this case and thus cannot confidently determine the clonality of these mutations.

We also identified an IPMN with low-grade dysplasia (IP12) that harbored three different inactivating *ARID1A* mutations in addition to a clonal *KRAS* mutation (Figure 3.4A). One mutation (p.A826Efs) occurred in 83% of single cells. The other two *ARID1A* mutations (p.P570Lfs and p.E1786Gfs) were mutually exclusive, occurring in in 25% and 13% of single cells, respectively, and were only present in cells with the first *ARID1A* alteration. Three other IPMNs IP16, IP27 (Figure 3.1), and IP08 (Figure 3.3B), also had subclonal *ARID1A* mutations, each present in only two to three single cells.

Mutations in other frequently altered driver genes in pancreatic ductal neoplasia occurred uncommonly in our cohort. Two IPMNs (IP08 and IP10) had clonal mutations in *CDKN2A*, while IP22 had a subclonal mutation in this gene. Although our cohort included four cases of IPMN with high-grade dysplasia and three cases of IPMN with associated adenocarcinoma, we identified *TP53* mutations in only one patient – unique *TP53* mutations were identified in IPMN and adjacent carcinoma in IP16. In addition, we only identified single mutations in *SMAD4* and in *TGFBR2*. Intriguingly, the *SMAD4* mutation co-occurred with an *RNF43* mutation in a subclone of IP22, which had only low-grade dysplasia, while the *TGFBR2* mutation occurred in the component IP16 with high-grade dysplasia.

Of note, in the single case of intraductal oncocytic papillary neoplasm (IOPN) (IP20), we did not identify any mutations in the genes in our panel, consistent with recent reports that these neoplasms are genetically distinct from other subtypes of IPMN[8].

Discussion

In this study, we provide the first single-cell genetic analysis of precursor lesions to invasive pancreatic cancer. In addition to demonstrating the feasibility of this type of analysis, these results provide insights into the genetic heterogeneity of early pancreatic tumorigenesis.

The majority of the IPMNs (seven of ten) had *KRAS* and/or *GNAS* mutations that were shared by the vast majority of analyzed cells. As the proportion of wild-type cells was similar to our ADO rate, the data suggest that these mutations are clonal in these IPMNs, that is, present in every neoplastic cell. In two of the ten IPMNs, we identified multiple clones with distinct *KRAS* mutations, suggesting the presence of independent neoplasms or neoplasms in which the shared genetic alteration does not occur in any of the known driver genes. In one IPMN (case IP16), the clone in the invasive carcinoma was genetically distinct from the two clones in the IPMN, consistent with an IPMN with concomitant rather than associated invasive adenocarcinoma[111]. The IPMNs in both IP16 and IP27 had multiple *KRAS* mutations within the same grossly defined IPMN. In these cases, the two clones were not grossly separated but instead were identified in the same small tissue fragment harvested from the wall of the IPMN. These data could suggest that these IPMNs were polyclonal, made up of multiple clones without a shared truncal genetic alteration. An alternative interpretation of these data is that these IPMNs

were monoclonal but initiated by an unidentified alteration prior to the development of *KRAS* mutations. The existence of such an earlier initiating alteration in pancreatic tumorigenesis has not yet been described.

The results of these studies also provide insight into the timing of driver gene mutations in pancreatic tumorigenesis. Like previous studies, our data suggest that mutations in *KRAS* and *GNAS* occur very early, as they are clonal in the majority of neoplasms in our study. In the IPMNs in which mutations in these genes were subclonal, there were no clonal mutations in other driver genes, suggesting that *KRAS* and *GNAS* are the earliest known driver genes in the pancreas. In contrast, mutations in *RNF43* and *ARID1A* are clearly subclonal in a subset of IPMNs, indicating that these mutations occur after the clonal *KRAS* or *GNAS* mutations. The persistence of cells without mutations in these genes highlights the heterogeneous clonal composition of IPMNs, perhaps suggesting that these driver gene mutations only provide slight selective advantage to certain cells within the tumor microenvironment. Of note, rare single cells (such as C_10 in IP10 and B_29 in IP24) lack mutations in *KRAS* and *GNAS* but have mutations in other driver genes. We interpret the lack of *KRAS/GNAS* mutations as a false negative result due to allelic drop out, a known artifact in single-cell sequencing data due to limited starting material. However, we cannot exclude that the identified genotypes are accurate, raising the possibility of a different sequence of driver gene mutations in a subset of cases.

In addition to mixtures of wild-type and mutant cells, our data also demonstrate that a subset of IPMNs consist of mixtures of neoplastic cells with different mutations in the same driver gene. In addition to clonal *KRAS* and *GNAS* mutations, IP22 also has

five different clones, each with unique mutations in *RNF43*. Similarly, IP24 has three unique mutually exclusive *RNF43* mutations. In contrast, IP12 has three different *ARID1A* mutations, two of which occurred as mutually exclusive second hits in small subclones of cells with the first *ARID1A* mutation. These IPMNs provide snapshots of the acquisition of tumor suppressor gene mutations in precursor lesions, suggesting a more complicated process than the sequential acquisition of two “hits”. Moreover, they suggest the presence of convergent evolution in at least a subset of IPMNs, in which mutations in a specific driver are strongly selected at a certain time point in tumorigenesis, resulting in multiple clones independently acquiring unique mutations in the selected gene. The identification of somatic mutations in *ARID1A* in IPMNs is also novel – although somatic mutations in this well-characterized tumor suppressor gene have been previously reported in PDAC, this is the first report of frequent *ARID1A* alterations in IPMNs[112].

A few technical considerations in our study are important to note. First, like Wang *et al*, we specifically isolated cells with replicated DNA (G2/M cells) in order to provide more template and thus improve the efficiency of our single-cell whole genome amplification[103]. Although this approach improves the technical success of our assay, selection of this subset does bias our analysis to proliferating cells. Even with this caveat, our data show that these cells represent a broad spectrum of clones with varying driver gene mutations. Comparison of mutation calls from the bulk and single-cell analyses also provides important insights into the utility of detailed single-cell analysis. Some of the subclonal mutations in *RNF43* in IP22 and IP24 were absent from the bulk sequencing data for the respective tissue fragments, despite sequencing at an

average coverage of almost 700X in bulk samples. Although our data suggest that single-cell sequencing has an increased sensitivity to detect rare clones, it is possible that ultra-deep sequencing (>1000X) of bulk tissue could achieve a similar sensitivity. Still, an ultra-deep bulk approach cannot assign rare mutations with similar frequency to specific clones, which is the true strength of single-cell analyses. There was also a mutation that was identified in a bulk sample but not in any single cells from that section (p.V31Dfs in *RNF43* in section A of IP22). This mutation had very low variant allele frequency in the bulk sample, suggesting that the number of single cells analyzed was likely not adequate to identify the rare cells with this mutation; such sensitivity issues will decrease as technical improvements allow analysis of larger numbers of single cells per sample. Overall, these findings highlight the strength of paired single-cell and bulk analysis of tissue samples, as the approaches are complementary. Finally, our data highlight the importance of assays to evaluate the quality of single-cell DNA amplification. Through our two-step filtering procedure (multiplex PCR, analysis of heterozygous germline SNPs), we restricted our analyses to only the most robustly amplified cells. Although this filtering excluded a significant proportion of the initially sorted cells (57%), it provides confidence in the quality of the data that passed these rigorous filters. In particular, we report a low ADO of 13% in analyzed cells, likely due to our analysis of only the most robustly amplified cells. Moreover, our FPR is consistent with those reported in other single-cell studies, and none of our control samples had false positive mutations that occurred in more than one single cell. This strongly suggests that even the rare subclones identified in our samples are unlikely to be artifacts caused by the extensive amplification required for single-cell sequencing.

Single-cell mutation calling remains an active research area in computational genomics, and the pipeline developed for this data set is novel. We found that currently available protocols to call mutations were not sufficiently optimized to take advantage of our sequencing data, which included IPMN single-cell samples, IPMN bulk samples, and matched normal bulk samples for each case in our study. The analysis pipeline developed in this work combines a standard variant caller designed for high specificity in bulk tissues and enhancements to handle single-cell amplification bias and increase caller sensitivity through multi-sample information pooling. The final set of mutation calls were a product of multiple tools and empirically selected thresholds; slight variations of these thresholds yielded stable results. We also utilized imputation to infer genotypes of cells for which sequencing data was indeterminate. Of note, imputation was not used in our study to identify new mutations but only to resolve indeterminate calls of mutations that had already been convincingly identified in other cells. Thus, this algorithm could only change the proportion of cells within an IPMN with a particular mutation.

Overall, our data provide the first insights into genetic heterogeneity of pancreatic cancer precursors at the single-cell level. Because our study encompassed a limited sample size of ten IPMNs, it is not possible to definitively determine the prevalence of this heterogeneity in patients with IPMN. Analysis of more cells and more lesions will be required to systematically catalogue this genetic heterogeneity and to correlate it with clinical features, such as grade of dysplasia and risk of malignant progression. Moreover, most of the IPMNs analyzed in our study were gastric-type, so our study provides limited insights into heterogeneity in other IPMN subtypes – all-inclusive cohorts with broad representation of all histological subtypes will be critical in future

studies to comprehensively describe the nature and extent of this single-cell genetic heterogeneity in IPMNs. Still, our studies suggest complex patterns of clonal evolution in preinvasive lesions. In addition, more extensive sequencing (such as whole exome sequencing) and identification of different types of alterations (such as copy number changes) will provide a more complete picture of clonal evolution in IPMNs, but our driver-focused approach provides key insights into heterogeneity of alterations that drive pancreatic tumorigenesis. Genetic heterogeneity with respect to these driver genes is most likely to have functional consequences in the heterogeneous clones and thus is likely to have biological importance, providing novel insights into pancreatic tumorigenesis.

Table 3.1. Clinicopathological and molecular data from analyzed IPMNs

Case	Sex	Age	Location	Histological Subtype	Final Diagnosis	Tissue Analyzed	Sections	Mutations identified
IP04	M	76	Head	Intestinal	IPMN with HGD	IPMN	A: IPMN F: IPMN G: IPMN	GNAS(p.R201C)
IP08	F	64	Tail	Gastric	IPMN with carcinoma (ductal)	IPMN, carcinoma	A: carcinoma B: IPMN C: IPMN	KRAS(p.G12V), CDKN2A(p.G23Vfs), ARID1A(p.V1825L)
IP10	F	65	Body	Intestinal	IPMN with carcinoma (colloid)*	IPMN	B: IPMN C: IPMN	GNAS(p.R201C), RNF43(p.L12Dfs), CDKN2A(p.G55Afs)
IP11	M	78	Head	Gastric	IPMN with LGD+	IPMN	A: IPMN	KRAS(p.G12D), GNAS(p.R201C)
IP12	M	59	Tail	Gastric	IPMN with LGD	IPMN	B: IPMN C: IPMN	KRAS(p.G12D), ARID1A(p.P70Lfs, p.A826Efs, p.E1786Gfs)
IP16	F	82	Head	Gastric	IPMN with carcinoma (ductal)	IPMN, carcinoma	A: IPMN B: IPMN C: carcinoma	KRAS(p.G12D, p.G12V, p.G12R), TP53(p.175H, p.C229*), ARID1A(p.T1031I), TGFBR2(p.R232Q),
IP20	F	85	Tail	Oncocytic	IPMN with HGD	IPMN	A: IPMN B: IPMN C: IPMN D: IPMN	None
IP22	M	65	Tail	Gastric	IPMN with LGD	IPMN	A: IPMN B: IPMN	KRAS(p.G12V), GNAS(p.R201C), RNF43(p.G3Sfs, p.V31Dfs, p.C91Y, p.G166V, p.VD299VY, p.H420Pfs), SMAD4 (p.E205K), CDKN2A(p.H83D)
IP24	F	63	Head	Gastric	IPMN with HGD+	IPMN	A: IPMN B: IPMN	KRAS(p.G12V), GNAS(p.R201C), RNF43(p.A11Lfs, p.T28I, p.G207D)
IP27	M	67	Head	Gastric	IPMN with HGD	IPMN	A: IPMN B: IPMN	KRAS(p.G12D, p.G12R), GNAS(p.R201H), ARID1A(p.M841Ifs)

*The carcinoma was not identified at the time of specimen processing and was not sampled for this study.
+These IPMNs co-occurred with grossly distinct biliary carcinomas that were not sampled for this study

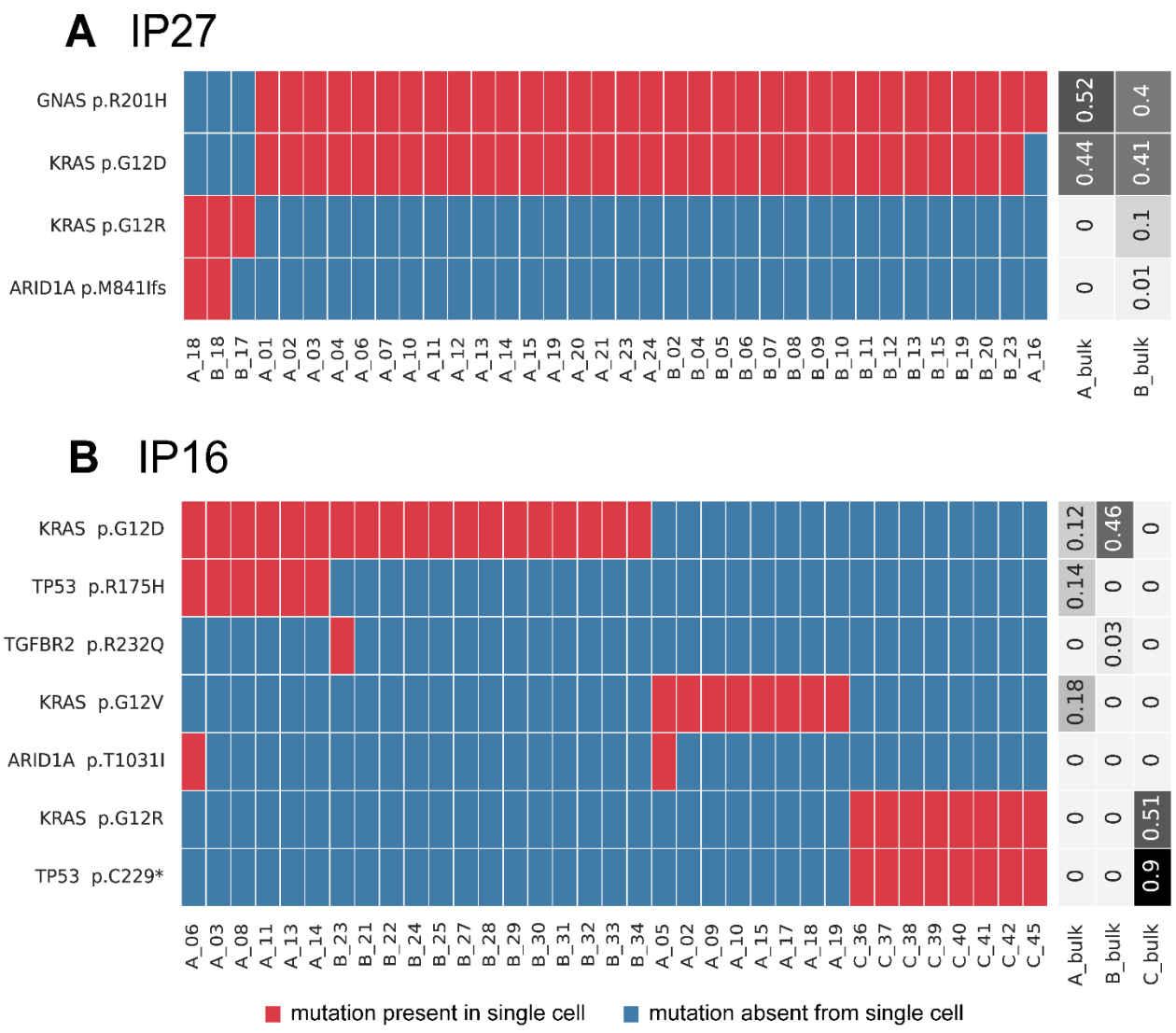

Figure 3.1. Somatic mutations identified in single cells from IPMNs with multiple *KRAS* mutations. Somatic mutations are presented in heatmaps with each row representing a mutation and each column representing a single cell. Single cells are designated by their tissue section (A-C) and cell number. Cells and mutations were clustered with Euclidean distance bi-clustering. The colors indicate the mutation calls after imputation, with red indicating mutant and blue indicating wild-type. Variant allele frequencies of the identified mutations in bulk samples from each section are indicated on the right. Both depicted IPMNs have multiple unique *KRAS* mutations. The majority of cells in IP27 (**A**), a gastric-type IPMN with high-grade dysplasia, have p.G12D in *KRAS* (as well as p.R201H in *GNAS*), while a small subclone lacks these mutations and instead has p.G12R in *KRAS*. IP16 (**B**) represents a gastric-type IPMN (sections A and B) with an adjacent ductal adenocarcinoma (section C). In this case, the IPMN contained two unique and mutually exclusive *KRAS* mutations, while the cancer had a third *KRAS* mutation as well as a unique mutation in *TP53*.

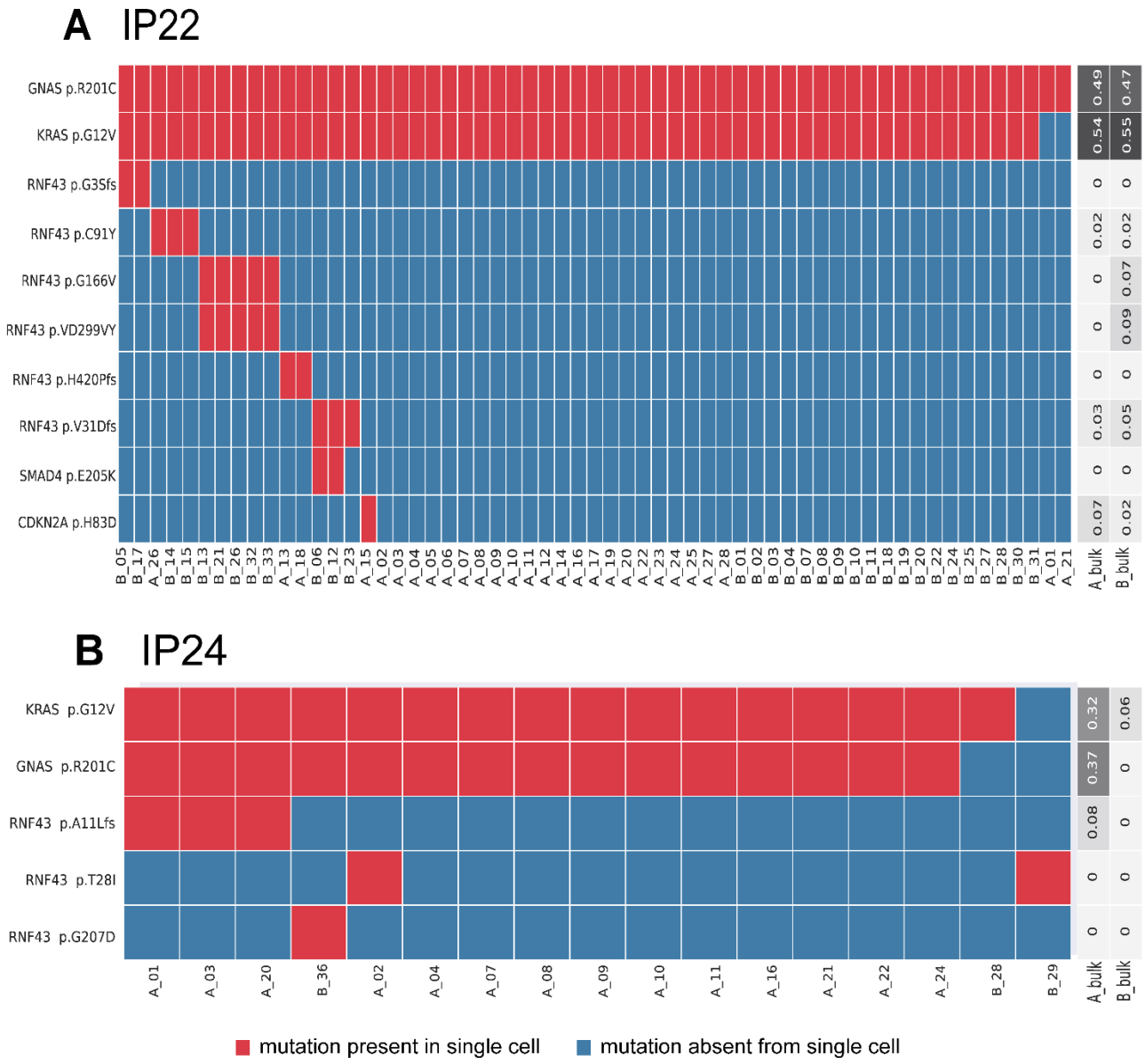

Figure 3.2. Somatic mutations identified in single cells from IPMNs with multiple *RNF43* mutations. Somatic mutations are presented in heatmaps with each row representing a mutation and each column representing a single cell. Single cells are designated by their tissue section (A-C) and cell number. Cells and mutations were clustered with Euclidean distance bi-clustering. The colors indicate the mutation calls after imputation, with red indicating mutant and blue indicating wild-type. Variant allele frequencies of the identified mutations in bulk samples from each section are indicated on the right. Both depicted IPMNs have multiple unique *RNF43* mutations, with three mutually exclusive *RNF43* mutations in IP24 (**A**) and five mutually exclusive subclones with different *RNF43* mutations in IP22 (**B**).

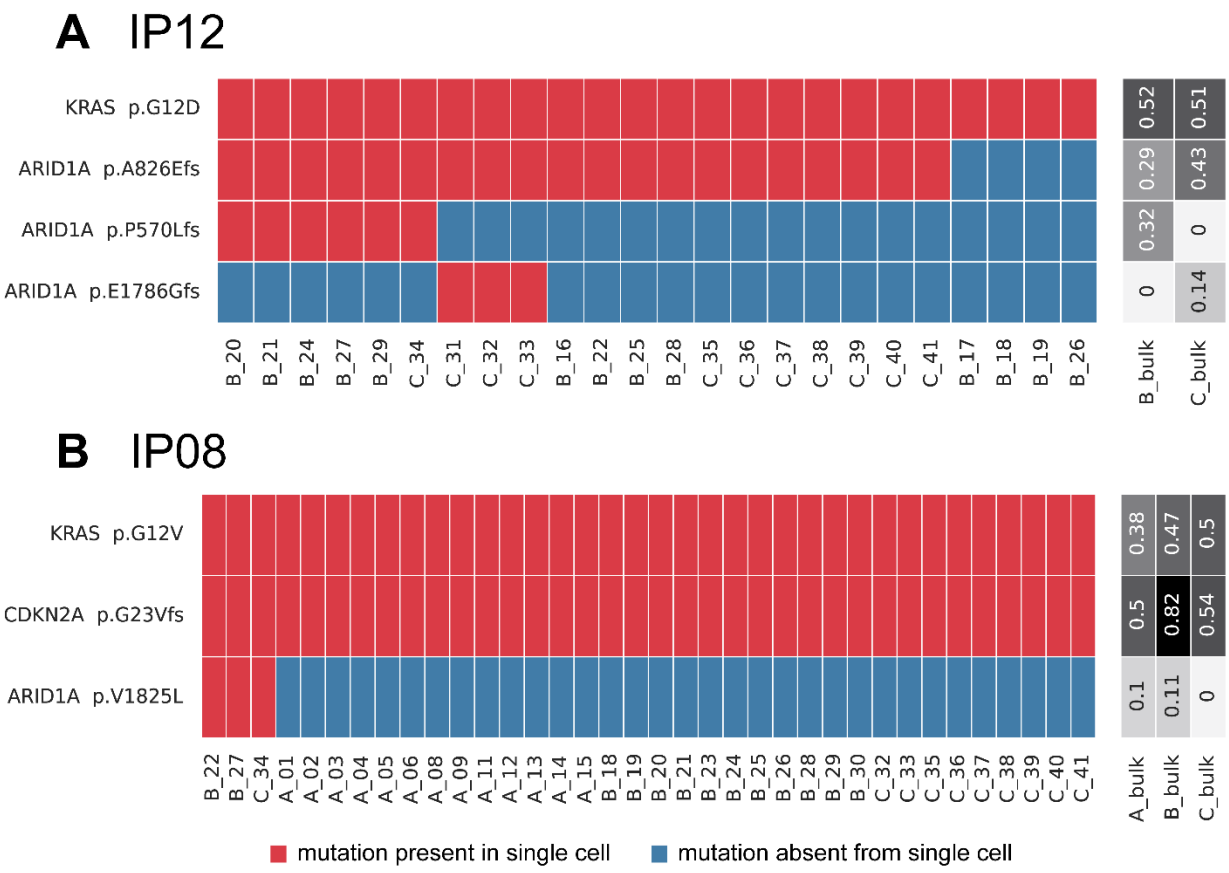

Figure 3.3. Somatic mutations identified in single cells from IPMNs with subclonal *ARID1A* mutations. Somatic mutations are presented in heatmaps with each row representing a mutation and each column representing a single cell. Single cells are designated by their tissue section (A-C) and cell number. Cells and mutations were clustered with Euclidean distance bi-clustering. The colors indicate the mutation calls after imputation, with red indicating mutant and blue indicating wild-type. Variant allele frequencies of the identified mutations in bulk samples from each section are indicated on the right. Both depicted IPMNs have subclonal *ARID1A* mutations. In IP12 (**A**), a gastric IPMN with low-grade dysplasia, there was a subclonal inactivating mutation representing a first hit in this tumor suppressor gene as well as two mutually exclusive second hits, each present in non-overlapping subclones. In IP08 (**B**), an IPMN with high-grade dysplasia, there is a single subclonal *ARID1A* mutation present in only three single cells.

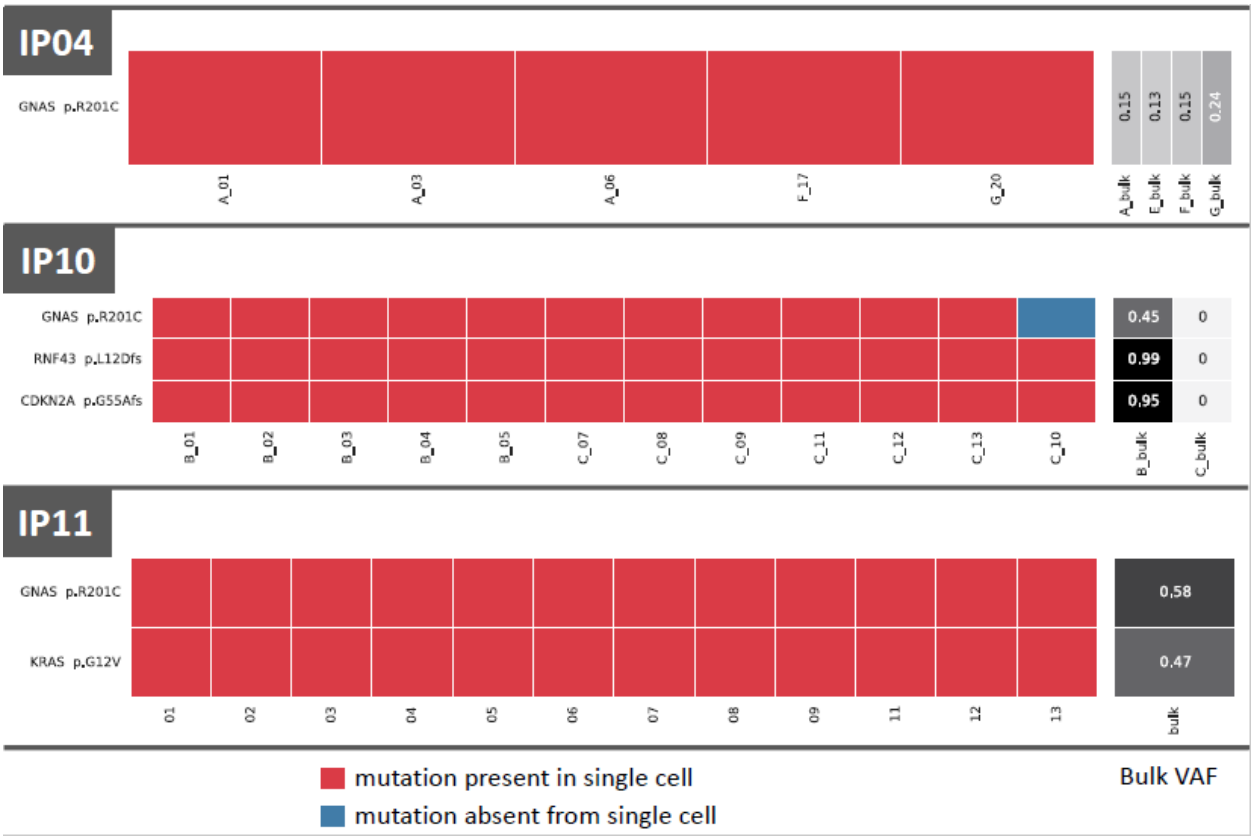

Figure 3.4. Somatic mutations identified in single cells from remaining cases. Somatic mutations are presented in heatmaps with each row representing a mutation and each column representing a single cell. Single cells are designated by their tissue section and/or cell number. Cells and mutations were clustered with Euclidean distance bi-clustering. The colors indicate the mutation calls after imputation, with red indicating mutant and blue indicating wild-type. Variant allele frequencies of the identified mutations in bulk samples from each section are indicated on the right. All depicted IPMNs have the driver mutation(s) present in every cell. In IP04, an IPMN with high-grade dysplasia, there is a *GNAS* p.R201C mutation. In IP10, an IPMN with high-grade dysplasia, there are clonal *GNAS* p.R201C, *RNF43* p.L12Dfs, and *CDKN2A* p.G55Afs mutations. In IP11, a gastric-type IPMN with low-grade dysplasia, there are clonal *GNAS* p.R201C and *KRAS* p.G12V mutations.

CHAPTER 4:

DISCUSSION

DISCUSSION

This work represents some of the first data describing the polyclonal origin of pancreatic precancers. The results of these studies have broadened our knowledge regarding somatic driver gene heterogeneity in pancreatic neoplasia, which have substantial implications for current early detection approaches and for our current understanding of the origins of pancreatic tumorigenesis.

Implications for field cancerization in the pancreas

A cancerized field is generally defined as a group of cells that are considered to be further along an evolutionary path towards cancer, but do not possess all the traits required for malignancy, and thus incapable of growing into a tumor[113,114]. The concept of field cancerization was first proposed by Slaughter *et al.* after observing microscopically abnormal epithelial fields adjacent to and, in some cases, immediately surrounding areas of oral squamous carcinoma[115]. Several decades later, improved molecular biological techniques and the advent of genetic sequencing have provided strong evidence for field cancerization in many tumor types. For example, a recent study found somatic driver gene alterations in normal airway epithelia, which were also shared with their matched non-small cell lung cancer (NSCLC), suggesting these variants were present prior to tumor formation and underwent clonal expansion during tumorigenesis[116]. In another study, Tao *et al.* demonstrated that spontaneous promoter hypermethylation, leading to epigenetic silencing, creates a permissive environment for oncogenic alterations (i.e. *BRAF* p.V600E) in a mouse-derived colon organoid model[117].

To date, there has been much speculation regarding field cancerization in pancreas. Patients diagnosed with IPMN often present with multiple lesions – as such, many have proposed this multifocality supports the concept of field cancerization. Furthermore, several studies have utilized genetic sequencing to determine the "relatedness" of these multifocal tumors. Matthaei *et al.* performed pyrosequencing of *KRAS*, coupled with LOH analysis of chromosomes 6q and 17p, to determine that the majority (~70%) of multifocal IPMNs are 'likely independent'[118]. In addition, Pea *et al.* used targeted NGS on IPMN and PDAC lesions from 13 patients who developed disease progression in their remnant pancreas following resection of IPMN[119]. Analysis of both the primary and recurrent IPMN revealed that more than half of the cases had genetic alterations not shared between these neoplasms and were classified as 'likely independent.' Finally, Felsenstein *et al.* described the genetic relationship between invasive carcinoma and co-occurring IPMN[22]. Interestingly, 18% of PDACs and co-occurring IPMNs did not share any driver gene mutation and were considered 'likely independent,' despite their close anatomic proximity. Altogether, these studies support the concept of field cancerization in a subset of patients, which leads to an increased risk of neoplasia throughout the pancreas.

The work described in Chapters 2 and 3 of this thesis provides further support for field cancerization, demonstrating that even a single IPMN lesion may arise from multiple, genetically independent clones. While the mechanism(s) driving field cancerization in the pancreas remains unknown, our WES data from Chapter 2 suggests that there is not a DNA alteration preceding *KRAS* that initiates tumorigenesis; however, we cannot completely rule out the possibility that a shared alteration occurred

(i.e. epigenetic alteration, whole exon loss, lariat base alteration, structural alteration, enhancer/promoter alteration), immediately followed by *KRAS* mutations. Polyclonal origin of precancerous lesions in other organ systems has previously been demonstrated, including esophagus and skin[77,78,120]. Intriguingly, many of these organs sites have known strong carcinogen exposures which may serve as a potential mechanism underlying polyclonal tumor origin. It is possible that an as-yet unidentified environmental exposure in the pancreas serves as the field carcinogen (i.e. focal pancreatitis, duodenal reflux, bile acids) leading to increased DNA mutations and emergence of clones with *KRAS* and *GNAS* mutations. Alternatively, some of the most recent data regarding field cancerization in pancreas provides a link between cellular metabolism and propensity for *KRAS* mutations[121]. This study found that hyperglycemic conditions in pancreatic cells leads to elevated cellular O-GlcNAcylation, which triggers genomic instability that preferentially increases *KRAS* mutations. Another recent study found that abundance of cellular acetyl-CoA plays a key role in tumor formation by regulating histone acetylation and mediating acinar-to-ductal metaplasia[122]. Importantly, however, the investigators found that this pathway is mutant *KRAS*-mediated, and does not function in *KRAS* WT cells. This finding suggest there is another event that demarcates the emergence of precancerous lesions from the normal-appearing cancerized field. It is possible that certain cellular metabolic processes cooperate by first, inducing the acquisition of *KRAS* mutations in normal pancreatic cells, which then contribute to additional metabolic reprogramming that promotes tumorigenesis. Both of the aforementioned studies predominately utilized 2D pancreatic cell cultures and/or genetically engineered mouse models (GEMMs). Future

studies should determine if *KRAS* mutations are present in normal human pancreatic tissues (i.e. ductal or acinar cells) in various disease-settings, including IPMN.

Implications for subclonal interactions in polyclonal tumors

The work described in Chapters 2 and 3 demonstrates that pancreatic precancers are polyclonal in origin, and we assume that some selective pressures lead to clonal dominance and expansion during tumor progression. Alternatively, pancreatic tumor growth may actually be supported by cross-talk between neoplastic cell subpopulations. Current evidence suggests that clonal evolution is not a stochastic process, but rather influenced by sublconal communication, such as paracrine signaling and cell-to-cell interactions[123]. For example, one study found that co-culturing distinct neoplastic subpopulations of small cell lung cancer (SCLC) led to enhanced tumor growth and, intriguingly, metastatic potential[124]. Others have demonstrated subclonal cooperation in breast cancer by utilizing GEMMs to show tumor growth is driven by subclones of cells which secrete tumor-promoting factors, such as IL11 and Wnt1, to stimulate expansion of other clones[125,126]. It is possible that similar mechanisms of clonal interactions are driving the progression of pancreatic precancers – as such, future studies should investigate the role of multiple *GNAS* and/or *KRAS* mutant clones in pancreatic tumor growth and malignant progression.

Conclusions

This thesis has provided strong evidence for the polyclonal origin of IPMNs, and transforms our understanding of the earliest stages of pancreatic tumorigenesis. In addition, these data reveal the substantial degree of driver gene heterogeneity within IPMNs, which has important clinical implications, particularly for current early detection

strategies. Finally, this work provides strong evidence supporting the concept of field cancerization in the pancreas. Discovering the mechanism(s) of this field cancerization will be critical for identifying individuals at highest risk for developing pancreatic neoplasia. This discovery has the potential to revolutionize the way clinicians screen and treat patients, and provide hope for those suffering from pancreatic cancer.

REFERENCES

1 Rahib L, Smith BD, Aizenberg R, *et al.* Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res* May 2014

2 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* 2018; **68**: 7-30

3 Laffan TA, Horton KM, Klein AP, *et al.* Prevalence of Unsuspected Pancreatic Cysts on MDCT. *AJR Am J Roentgenol* 2008; **191**: 802-807

4 Lee KS, Sekhar A, Rofsky NM, *et al.* Prevalence of Incidental Pancreatic Cysts in the Adult Population on MR Imaging. *The American Journal of Gastroenterology* 2010; **105**: 2079-2084

5 Hruban RH, Takaori K, Klimstra DS, *et al.* An Illustrated Consensus on the Classification of Pancreatic Intraepithelial Neoplasia and Intraductal Papillary Mucinous Neoplasms. *The American Journal of Surgical Pathology* 2004; **28**: 977-987

6 Basturk O, Hong S-M, Wood LD, *et al.* A REVISED CLASSIFICATION SYSTEM AND RECOMMENDATIONS FROM THE BALTIMORE CONSENSUS MEETING FOR NEOPLASTIC PRECURSOR LESIONS IN THE PANCREAS. *Am J Surg Pathol* 2015; **39**: 1730-1741

7 Basturk O, Chung SM, Hruban RH, *et al.* Distinct Pathways of Pathogenesis of Intraductal Oncocytic Papillary Neoplasms and Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Virchows Arch* 2016; **469**: 523-532

8 Basturk O, Tan M, Bhanot U, *et al.* The oncocytic subtype is genetically distinct from other pancreatic intraductal papillary mucinous neoplasm subtypes. *Mod Pathol* 2016; **29**: 1058-1069

9 Adsay NV, Merati K, Basturk O, *et al.* Pathologically and Biologically Distinct Types of Epithelium in Intraductal Papillary Mucinous Neoplasms: Delineation of an ‘intestinal’ Pathway of Carcinogenesis in the Pancreas. *The American Journal of Surgical Pathology* 2004; **28**: 839-848

10 Adsay NV, Pierson C, Sarkar F, *et al.* Colloid (mucinous Noncystic) Carcinoma of the Pancreas. *The American Journal of Surgical Pathology* 2001; **25**: 26-42

11 Grützmann R, Niedergethmann M, Pilarsky C, *et al.* Intraductal Papillary Mucinous Tumors of the Pancreas: Biology, Diagnosis, and Treatment. *The Oncologist* 2010; **15**: 1294-1309

12 Sohn TA, Yeo CJ, Cameron JL, *et al.* Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Ann Surg* 2004; **239**: 788-799

13 Tanaka M, Fernández-del Castillo C, Adsay V, *et al.* International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. *Pancreatology* 2012; **12**: 183-197

14 Reid MD, Saka B, Balci S, *et al.* Molecular Genetics of Pancreatic Neoplasms and Their Morphologic CorrelatesAn Update on Recent Advances and Potential Diagnostic Applications. *Am J Clin Pathol* 2014; **141**: 168-180

15 Wu J, Jiao Y, Molin MD, *et al.* Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *PNAS* 2011; **108**: 21188-21193

16 Kuboki Y, Shimizu K, Hatori T, *et al.* Molecular Biomarkers for Progression of Intraductal Papillary Mucinous Neoplasm of the Pancreas. *Pancreas* 2015; **44**: 227-235

17 Amato E, Molin M dal, Mafficini A, *et al.* Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas. *J Pathol* 2014; **233**: 217-227

18 Hosoda W, Sasaki E, Murakami Y, *et al.* GNAS mutation is a frequent event in pancreatic intraductal papillary mucinous neoplasms and associated adenocarcinomas. *Virchows Arch* 2015; **466**: 665-674

19 Wu J, Matthaei H, Maitra A, *et al.* Recurrent GNAS Mutations Define an Unexpected Pathway for Pancreatic Cyst Development. *Science Translational Medicine* 2011; **3**: 92ra66-ra92ra66

20 Witkiewicz AK, McMillan EA, Balaji U, *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature Communications* 2015; **6**: 6744

21 Furukawa T, Kuboki Y, Tanji E, *et al.* Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas. *Sci Rep* 2011; **1**

22 Felsenstein M, Noë M, Masica DL, *et al.* IPMNs with co-occurring invasive cancers: neighbours but not always relatives. *Gut* March 2018: gutjnl - 2017-315062

23 Lania AG, Mantovani G, Spada A. Mechanisms of Disease: mutations of G proteins and G-protein-coupled receptors in endocrine diseases. *Nature Reviews Endocrinology* 2006; **2**: 681-693

24 Tan MC, Basturk O, Brannon AR, *et al.* GNAS and KRAS Mutations Define Separate Progression Pathways in Intraductal Papillary Mucinous Neoplasm-Associated Carcinoma. *J Am Coll Surg* 2015; **220**: 845-854.e1

25 Molin MD, Matthaei H, Wu J, *et al.* Clinicopathological Correlates of Activating GNAS Mutations in Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas. *Ann Surg Oncol* 2013; **20**

26 Koo B-K, Spit M, Jordens I, *et al.* Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* 2012; **488**: 665-669

27 Waddell N, Pajic M, Patch A-M, *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015; **518**: 495-501

28 Molin MD, Zhang M, Wilde RF de, *et al.* Very Long-term Survival Following Resection for Pancreatic Cancer Is Not Explained by Commonly Mutated Genes: Results of Whole-Exome Sequencing Analysis. *Clin Cancer Res* 2015; **21**: 1944-1950

29 Biankin AV, Biankin SA, Kench JG, *et al.* Aberrant p16INK4A and DPC4/Smad4 expression in intraductal papillary mucinous tumours of the pancreas is associated with invasive ductal adenocarcinoma. *Gut* 2002; **50**: 861-868

30 Mohri D, Asaoka Y, Ijichi H, *et al.* Different subtypes of intraductal papillary mucinous neoplasm in the pancreas have distinct pathways to pancreatic cancer progression. *J Gastroenterol* 2012; **47**: 203-213

31 Sato N, Ueki T, Fukushima N, *et al.* Aberrant methylation of CpG islands in intraductal papillary mucinous neoplasms of the pancreas. *Gastroenterology* 2002; **123**: 365-372

32 Abe T, Fukushima N, Brune K, *et al.* Genome-Wide Allelotypes of Familial Pancreatic Adenocarcinomas and Familial and Sporadic Intraductal Papillary Mucinous Neoplasms. *Clin Cancer Res* 2007; **13**: 6019-6025

33 Fujii H, Inagaki M, Kasai S, *et al.* Genetic progression and heterogeneity in intraductal papillary-mucinous neoplasms of the pancreas. *American Journal of Pathology* 1997; **151**: 1447-1454

34 Chadwick B, Willmore-Payne C, Tripp S, *et al.* Histologic, immunohistochemical, and molecular classification of 52 IPMNs of the pancreas. *Appl Immunohistochem Mol Morphol* 2009; **17**: 31-39

35 Sasaki S, Yamamoto H, Kaneto H, *et al.* Differential roles of alterations of p53, p16, and SMAD4 expression in the progression of intraductal papillary-mucinous tumors of the pancreas. *Oncology Reports* 2003; **10**: 21-25

36 Satoh K, Shimosegawa T, Moriizumi S, *et al.* K-ras mutation and p53 protein accumulation in intraductal mucin-hypersecreting neoplasms of the pancreas. *Pancreas* 1996; **12**: 362-368

37 Iacobuzio-Donahue CA, Klimstra DS, Adsay NV, *et al.* Dpc-4 Protein Is Expressed in Virtually All Human Intraductal Papillary Mucinous Neoplasms of the Pancreas: Comparison with Conventional Ductal Adenocarcinomas. *The American Journal of Pathology* 2000; **157**: 755-761

38 Furukawa T, Fujisaki R, Yoshida Y, *et al.* Distinct progression pathways involving the dysfunction of DUSP6/MKP-3 in pancreatic intraepithelial neoplasia and intraductal papillary-mucinous neoplasms of the pancreas. *Modern Pathology* 2005; **18**: 1034-1042

39 Schönleben F, Qiu W, Ciau NT, *et al.* PIK3CA Mutations in Intraductal Papillary Mucinous Neoplasm/Carcinoma of the Pancreas. *Clin Cancer Res* 2006; **12**: 3851-3855

40 Schönleben F, Qiu W, Bruckman KC, *et al.* BRAF and KRAS gene mutations in intraductal papillary mucinous neoplasm/carcinoma (IPMN/IPMC) of the pancreas. *Cancer Letters* 2007; **249**: 242-248

41 Sato N, Rosty C, Jansen M, *et al.* STK11/LKB1 Peutz-Jeghers Gene Inactivation in Intraductal Papillary-Mucinous Neoplasms of the Pancreas. *The American Journal of Pathology* 2001; **159**: 2017-2022

42 Garcia-Carracedo D, Turk AT, Fine SA, *et al.* Loss of PTEN Expression Is Associated with Poor Prognosis in Patients with Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Cancer Res* 2013; **19**: 6830-6841

43 Durante S, Vecchiarelli S, Astolfi A, *et al.* Copy number gain of chromosome 3q is a recurrent event in patients with intraductal papillary mucinous neoplasm (IPMN) associated with disease progression. *Oncotarget* 2016; **7**: 74797-74806

44 Kopp JL, Dubois CL, Schaeffer DF, *et al.* Loss of Pten and Activation of Kras Synergistically Induce Formation of Intraductal Papillary Mucinous Neoplasia From Pancreatic Ductal Cells in Mice. *Gastroenterology* 2018; **154**: 1509-1523.e5

45 Ying H, Elpek KG, Vinjamoori A, *et al.* Pten is a major tumor suppressor in pancreatic ductal adenocarcinoma and regulates an NF-κB-cytokine network. *Cancer Discov* 2011; **1**: 158-169

46 Hemminki A, Markie D, Tomlinson I, *et al.* A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature* 1998; **391**: 184-187

47 Boardman LA. Increased Risk for Cancer in Patients with the Peutz-Jeghers Syndrome. *Annals of Internal Medicine* 1998; **128**: 896

48 Resta N, Simone C, Mareni C, *et al.* STK11 mutations in Peutz-Jeghers syndrome and sporadic colon cancer. *Cancer Res* 1998; **58**: 4799-4801

49 Forster LF, Defres S, Goudie DR, *et al.* An investigation of the Peutz-Jeghers gene (LKB1) in sporadic breast and colon cancers. *J Clin Pathol* 2000; **53**: 791-793

50 Wang ZJ, Churchman M, Campbell IG, *et al.* Allele loss and mutation screen at the Peutz-Jeghers (LKB1) locus (19p13.3) in sporadic ovarian tumours. *Br J Cancer* 1999; **80**: 70-72

51 Su GH, Hruban RH, Bansal RK, *et al.* Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. *Am J Pathol* 1999; **154**: 1835-1840

52 Canto MI, Hruban RH, Fishman EK, *et al.* Frequent Detection of Pancreatic Lesions in Asymptomatic High-Risk Individuals. *Gastroenterology* 2012; **142**: 796-804

53 Neesse A, Hahnenkamp A, Griesmann H, *et al.* Claudin-4-targeted optical imaging detects pancreatic cancer and its precursor lesions. *Gut* 2013; **62**: 1034-1043

54 Gress TM, Müller-Pillasch F, Geng M, *et al.* A pancreatic cancer-specific expression profile. *Oncogene* 1996; **13**: 1819-1830

55 Geng MM, Ellenrieder V, Wallrapp C, *et al.* Use of representational difference analysis to study the effect of TGFB on the expression profile of a pancreatic cancer cell line. *Genes Chromosomes Cancer* 1999; **26**: 70-79

56 Genevay M, Mino-Kenudson M, Yaeger K, *et al.* Cytology Adds Value to Imaging Studies for Risk Assessment of Malignancy in Pancreatic Mucinous Cysts. *Ann Surg* 2011; **254**

57 Schmidt CM, White PB, Waters JA, *et al.* Intraductal Papillary Mucinous Neoplasms: Predictors of Malignant and Invasive Pathology. *Annals of Surgery* 2007; **246**: 644

58 Wiesenauer CA, Schmidt CM, Cummings OW, *et al.* Preoperative Predictors of Malignancy in Pancreatic Intraductal Papillary Mucinous Neoplasms. *Arch Surg* 2003; **138**: 610-618

59 Allen PJ, Qin L-X, Tang L, *et al.* Pancreatic Cyst Fluid Protein Expression Profiling for Discriminating Between Serous Cystadenoma and Intraductal Papillary Mucinous Neoplasm. *Ann Surg* 2009; **250**: 754-760

60 Lewandrowski KB, Southern JF, Pins MR, *et al.* Cyst fluid analysis in the differential diagnosis of pancreatic cysts. A comparison of pseudocysts, serous cystadenomas, mucinous cystic neoplasms, and mucinous cystadenocarcinoma. *Ann Surg* 1993; **217**: 41-47

61 Brugge WR, Lewandrowski K, Lee-Lewandrowski E, *et al.* Diagnosis of pancreatic cystic neoplasms: a report of the cooperative pancreatic cyst study. *Gastroenterology* 2004; **126**: 1330-1336

62 Singhi AD, McGrath K, Brand RE, *et al.* Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia. *Gut* September 2017: gutjnl - 2016-313586

63 Springer S, Wang Y, Molin MD, *et al.* A Combination of Molecular Markers and Clinical Features Improve the Classification of Pancreatic Cysts. *Gastroenterology* 2015; **149**: 1501-1510

64 Hata T, Dal Molin M, McGregor-Das A, *et al.* Simple Detection of Telomere Fusions in Pancreatic Cancer, Intraductal Papillary Mucinous Neoplasm, and Pancreatic Cyst Fluid. *The Journal of Molecular Diagnostics* 2018; **20**: 46-55

65 Hata T, Molin MD, Suenaga M, *et al.* Cyst Fluid Telomerase Activity Predicts the Histologic Grade of Cystic Neoplasms of the Pancreas. *Clin Cancer Res* 2016; **22**: 5141-5151

66 Ryu JK, Matthaei H, dal Molin M, *et al.* Elevated microRNA miR-21 Levels in Pancreatic Cyst Fluid Are Predictive of Mucinous Precursor Lesions of Ductal Adenocarcinoma. *Pancreatology* 2011; **11**: 343-350

67 Yu J, Sadakari Y, Shindo K, *et al.* Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms. *Gut* 2017; **66**: 1677-1687

68 Suenaga M, Yu J, Shindo K, *et al.* Pancreatic juice mutation concentrations can help predict the grade of dysplasia in patients undergoing pancreatic surveillance. *Clin Cancer Res* January 2018: clincanres.2463.2017

69 Poruk KE, Valero V, He J, *et al.* Circulating Epithelial Cells in Intraductal Papillary Mucinous Neoplasms and Cystic Pancreatic Lesions. *Pancreas* 2017; **46**: 943-947

70 Rhim AD, Thege FI, Santana SM, *et al.* Detection of Circulating Pancreas Epithelial Cells in Patients with Pancreatic Cystic Lesions. *Gastroenterology* 2014; **146**: 647-651

71 Young MR, Wagner PD, Ghosh S, *et al.* Validation of Biomarkers for Early Detection of Pancreatic Cancer: Summary of The Alliance of Pancreatic Cancer Consortia for Biomarkers for Early Detection Workshop. *Pancreas* 2018; **47**: 135-141

72 Fischer CG, Wood LD. From somatic mutation to early detection: insights from molecular characterization of pancreatic cancer precursor lesions. *The Journal of Pathology* 2018; **246**: 395-404

73 Felsenstein M, Hruban RH, Wood LD. New Developments in the Molecular Mechanisms of Pancreatic Tumorigenesis. *Advances in Anatomic Pathology* 2018; **25**: 131

74 Omori Y, Ono Y, Tanino M, *et al.* Pathways of Progression From Intraductal Papillary Mucinous Neoplasm to Pancreatic Ductal Adenocarcinoma Based on Molecular Features. *Gastroenterology* October 2018

75 Connor AA, Denroche RE, Jang GH, *et al.* Integration of Genomic and Transcriptional Features in Pancreatic Cancer Reveals Increased Cell Cycle Progression in Metastases. *Cancer Cell* 2019; **35**: 267-282.e7

76 Makohon-Moore AP, Zhang M, Reiter JG, *et al.* Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet* 2017; **49**: 358-366

77 Ross-Innes CS, Becq J, Warren A, *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nature Genetics* 2015; **47**: 1038-1046

78 Stachler MD, Taylor-Weiner A, Peng S, *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nature Genetics* 2015; **47**: 1047-1055

79 McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-1303

80 Van der Auwera GA, Carneiro MO, Hartl C, *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; **11**: 11.10.1-11.10.33

81 Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 2010; **26**: 589-595

82 Picard Tools - By Broad Institute. [Accessed February 20, 2019] Available from: http://broadinstitute.github.io/picard/.

83 Cibulskis K, Lawrence MS, Carter SL, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213-219

84 Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.* Integrative Genomics Viewer. *Nat Biotechnol* 2011; **29**: 24-26

85 Lin M-T, Mosier SL, Thiess M, *et al.* Clinical Validation of KRAS, BRAF, and EGFR Mutation Detection Using Next-Generation Sequencing. *Am J Clin Pathol* 2014; **141**: 856-866

86 Niknafs N, Beleva-Guthrie V, Naiman DQ, *et al.* SubClonal Hierarchy Inference from Somatic Mutations: Automatic Reconstruction of Cancer Evolutionary Trees from Multi-region Next Generation Sequencing. *PLoS Comput Biol* 2015; **11**

87 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114-2120

88 Saunders CT, Wong WSW, Swamy S, *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* 2012; **28**: 1811-1817

89 Wang F, Flanagan J, Su N, *et al.* RNAscope. *J Mol Diagn* 2012; **14**: 22-29

90 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

91 Chang MT, Asthana S, Gao SP, *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016; **34**: 155-163

92 Zehir A, Benayed R, Shah RH, *et al.* Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. *Nat Med* 2017; **23**: 703-713

93 Chen S-H, Zhang Y, Van Horn RD, *et al.* Oncogenic BRAF Deletions That Function as Homodimers and Are Sensitive to Inhibition by RAF Dimer Inhibitor LY3009120. *Cancer Discov* 2016; **6**: 300-315

94 Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 2017; **32**: 185-203.e13

95 Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* 2013; **110**: 1999-2004

96 Gerlinger M, Horswell S, Larkin J, *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; **46**: 225-233

97 Misale S, Nicolantonio FD, Sartore-Bianchi A, *et al.* Resistance to Anti-EGFR Therapy in Colorectal Cancer: From Heterogeneity to Convergent Evolution. *Cancer Discov* 2014; **4**: 1269-1280

98 Singhi AD, Nikiforova MN, Fasanella KE, *et al.* Preoperative GNAS and KRAS Testing in the Diagnosis of Pancreatic Mucinous Cysts. *Clin Cancer Res* 2014; **20**: 4381-4389

99 Pittman ME, Rao R, Hruban RH. Classification, Morphology, Molecular Pathogenesis, and Outcome of Premalignant Lesions of the Pancreas. *Archives of Pathology & Laboratory Medicine* 2017; **141**: 1606-1614

100 Tanaka M, Fernández-del Castillo C, Adsay V, *et al.* International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. *Pancreatology* 2012; **12**: 183-197

101 Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc Natl Acad Sci U S A* 2014; **111**: 17947-17952

102 Leung ML, Davis A, Gao R, *et al.* Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res* 2017; **27**: 1287-1299

103 Wang Y, Waters J, Leung ML, *et al.* Clonal Evolution in Breast Cancer Revealed by Single Nucleus Genome Sequencing. *Nature* 2014; **512**: 155-160

104 Xu X, Hou Y, Yin X, *et al.* Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor. *Cell* 2012; **148**: 886-895

105 de Bourcy CFA, De Vlaminck I, Kanbar JN, *et al.* A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *PLoS One* 2014; **9**

106 Hou Y, Wu K, Shi X, *et al.* Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *Gigascience* 2015; **4**

107 Hou Y, Song L, Zhu P, *et al.* Single-Cell Exome Sequencing and Monoclonal Evolution of a JAK2-Negative Myeloproliferative Neoplasm. *Cell* 2012; **148**: 873-885

108 Rago C, Huso DL, Diehl F, *et al.* Serial Assessment of Human Tumor Burdens in Mice by the Analysis of Circulating DNA. *Cancer Res* 2007; **67**: 9364-9370

109 [1109.2378] Modern hierarchical, agglomerative clustering algorithms. [Accessed April 25, 2019] Available from: https://arxiv.org/abs/1109.2378#.

110 DePristo MA, Banks E, Poplin RE, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-498

111 Yamaguchi K, Kanemitsu S, Hatori T, *et al.* Pancreatic Ductal Adenocarcinoma Derived From IPMN and Pancreatic Ductal Adenocarcinoma Concomitant With IPMN. *Pancreas* 2011; **40**: 571

112 Sausen M, Phallen J, Adleff V, *et al.* Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. *Nat Commun* 2015; **6**: 7686

113 Braakhuis BJM, Tabor MP, Kummer JA, *et al.* A Genetic Explanation of Slaughter's Concept of Field Cancerization: Evidence and Clinical Implications. *Cancer Res* 2003; **63**: 1727-1730

114 Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nature Reviews Cancer* 2018; **18**: 19-32

115 Slaughter DP, Southwick HW, Smejkal W. 'Field cancerization' in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer* 1953; **6**: 963-968

116 Kadara H, Sivakumar S, Jakubek Y, *et al.* Driver Mutations in Normal Airway Epithelium Elucidate Spatiotemporal Resolution of Lung Cancer. *Am J Respir Crit Care Med* March 2019

117 Tao Y, Kang B, Petkovich DA, *et al.* Aging-like Spontaneous Epigenetic Silencing Facilitates Wnt Activation, Stemness, and BrafV600E-Induced Tumorigenesis. *Cancer Cell* 2019; **35**: 315-328.e6

118 Matthaei H, Norris AL, Tsiatis AC, *et al.* Clinicopathological Characteristics and Molecular Analyses of Multifocal Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Ann Surg* 2012; **255**: 326-333

119 Pea A, Yu J, Rezaee N, *et al.* Targeted DNA Sequencing Reveals Patterns of Local Progression in the Pancreatic Remnant Following Resection of Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas. *Ann Surg* 2017; **266**: 133-141

120 Martincorena I, Roshan A, Gerstung M, *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015; **348**: 880-886

121 Hu C-M, Tien S-C, Hsieh P-K, *et al.* High Glucose Triggers Nucleotide Imbalance through O-GlcNAcylation of Key Enzymes and Induces KRAS Mutation in Pancreatic Cells. *Cell Metabolism* March 2019

122 Carrer A, Trefely S, Zhao S, *et al.* Acetyl-CoA Metabolism Supports Multistep Pancreatic Tumorigenesis. *Cancer Discov* 2019; **9**: 416-435

123 Gupta RG, Somer RA. Intratumor Heterogeneity: Novel Approaches for Resolving Genomic Architecture and Clonal Evolution. *Mol Cancer Res* 2017; **15**: 1127-1137

124 Calbo J, van Montfort E, Proost N, *et al.* A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. *Cancer Cell* 2011; **19**: 244-256

125 Cleary AS, Leonard TL, Gestl SA, *et al.* Tumor cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* 2014; **508**: 113-117

126 Marusyk A, Tabassum DP, Altrock PM, *et al.* Non-cell autonomous tumor-growth driving supports sub-clonal heterogeneity. *Nature* 2014; **514**: 54-58

CURRICULUM VITAE

Catherine G. Fischer

PERSONAL INFORMATION

19 N. Highland Ave Baltimore, MD 21224
412-657-6918
cmg5449@gmail.com or cguerra2@jhmi.edu
Borm April 13, 1992 in Pittsburgh, PA, USA

EDUCATION

Ph.D in Pathobiology **08/2014-present**
The Johns Hopkins University School of Medicine
Department of Pathology
Baltimore, MD, USA

B.S. with honors in Veterinary and Biomedical Sciences **08/2010-05/2014**
Minor in Biochemistry and Molecular Biology
The Pennsylvania State University
Schreyer Honors College, Department of Agricultural Sciences
University Park, PA, USA
Cumulative GPA: 3.89/4.00

RESEARCH EXPERIENCE

Doctoral Candidate **08/2014-present**
The Johns Hopkins University School of Medicine
Baltimore, MD
Research Advisor: Dr. Laura Wood
Project: Characterizing driver gene heterogeneity and clonal origin of intraductal papillary mucinous neoplasms of the pancreas

Undergraduate Honors Thesis Research **08/2012-05/2014**
The Pennsylvania State University
University Park, PA
Research Advisor: Dr. Jason Brooks
Project: *Fusobacterium varium* Infection in Mice as a Model for the Study of Vaccine Efficacy and Immunogenicity

REGULATORY AFFAIRS EXPERIENCE

Regulatory Affairs Fellow **01/2018-08/2018**
Regulatory Affairs Professionals Society (RAPS)
The Johns Hopkins University School of Medicine
Baltimore, MD US
Supervisor: Dr. Patricia Phelps and Linda Bowen

LEADERSHIP EXPERIENCE

Graduate Student Coordinator **03/2016-present**
The Johns Hopkins University School of Medicine
Baltimore, MD
Supervisor: Vice Dean of Education, Dr. Roy Ziegelstein

Graduate Student Representative, University Health Committee **09/2015-02/2018**

The Johns Hopkins University School of Medicine
Baltimore, MD
Supervisor: Dr. Spyridon Marinopoulos

TEACHING EXPERIENCE

Teaching Assistant, Grant Writing 101 Course **02/2017-05/2017**
The Johns Hopkins University School of Medicine
Baltimore, MD
Supervisor: Dr. James Eshleman

Teaching Assistant, Pathobiology: Basic Mechanisms of Disease 08/2015-10/2015
The Johns Hopkins University School of Medicine
Baltimore, MD
Supervisor: Dr. Kathleen Gabrielson

HONORS AND AWARDS

1st Place Poster Award
09/2015
The Pathobiology Graduate Program Annual Retreat

Ronald E. McNair Post-Baccalaureate Achievement Program
08/2013-05/2014
The Pennsylvania State University

Enrollment Incentive Award Scholarship
08/2010-05/2014
The Pennsylvania State University College of Agriculture Sciences

PROFESSIONAL SOCIETY MEMBERSHIPS

American Association for Cancer Research (AACR)
06/2018-Present
Regulatory Affairs Professionals Society (RAPS)
03/2017-Present
Phi Kappa Phi National Honors Society
05/2014

SELECTED SCIENTIFIC PRESENTATIONS

"Next-generation sequencing defines striking genetic heterogeneity in intraductal papillary mucinous neoplasms" Abstract accepted for Poster Presentation. AACR Pancreatic Cancer: Advances in Science and Clinical Care. September 2018.

"Next-generation sequencing defines striking genetic heterogeneity in intraductal papillary mucinous neoplasms of the pancreas" Selected Senior Oral Presentation. Johns Hopkins School of Medicine, Pathobiology Graduate Program Annual Retreat. September 2018.

"Genetic Heterogeneity in Intraductal Papillary Mucinous Neoplasms of the Pancreas" Oral Presentation. Johns Hopkins School of Medicine, Pathology Grand Rounds. June 2016.

"*Fusobacterium varium* Infection in Mice as a Model for the Study of Vaccine Efficacy and Immunogenicity" Oral Presentation. The Pennsylvania State SROP and McNair Research Symposium. July 2013

BIBLIOGRAPHY

Fischer CG*, Beleva Guthrie V*, Braxton AM, Zheng L, Wang P, Song Q, Griffin JF, Chianchiano PE, Hosoda W, Niknafs N, Springer S, Dal Molin M, Masica D, Scharpf RB, Thompson ED, He J, Wolfgang CL, Hruban RH, Roberts NJ, Lennon AM, Jiao Y, Karchin R, Wood LD. Polyclonal origin defines early pancreatic tumorigenesis. *Manuscript in review to Gastroenterology.*
(* represents co-first authorship)

Fischer CG, Wood LD. From somatic mutation to early detection: insights from molecular characterization of pancreatic cancer precursor lesions. *The Journal of Pathology* PMID: 30105857.

Kuboki Y*, **Fischer CG*,** Beleva Guthrie V*, Huang W, Yu J, Chianchiano P, Hosoda W, Zhang H, Zheng L, Shao M, Thompson ED, Waters K, Poling J, He J, Weiss MJ, Wolfgang CL, Goggins MG, Hruban RH, Roberts NJ, Karchin R, Wood LD. Single-Cell

Sequencing Defines Genetic Heterogeneity in Pancreatic Cancer Precursor Lesions. *The Journal of Pathology.* PMID: 30430578
(* represents co-first authorship)

Felsenstein M, Noe M, Hosoda W, Chianchiano P, **Fischer CG**, Brosens LAA, Pea A, Yu J, Gemenetzis G, Groot V, Makary MA, He J, Weiss MJ, Cameron JL, Wolfgang CL, Hruban RH, Roberts NJ, Goggins MG, Wood LD. (2018) IPMNs with co-occurring invasive cancers: neighbors but not always relatives. *Gut* PMID: 29500184.

Ideno N, Yamaguchi H, Ghosh B, Gupta S, Okumura T, Steffen DJ, **Fischer CG**, Wood LD, Singhi AD, Nakamura M, Gutkind JS, Maitra A. (2018) $GNAS^{R210C}$ Induces Pancreatic Cystic Neoplasms in Mice That Express Activated KRAS by Inhibiting YAP1 Signaling. *Gastroenterology* PMID: 30142336.